

# Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes

Juan A. Cuesta-Albertos<sup>1</sup>

Eduardo García-Portugués<sup>2,4</sup>  
Wenceslao González-Manteiga<sup>3</sup>

Manuel Febrero-Bande<sup>3</sup>

## Abstract

We consider marked empirical processes, indexed by a randomly projected functional covariate, to construct goodness-of-fit tests for the functional linear model with scalar response. The test statistics are built from continuous functionals over the projected process, resulting in computationally efficient tests that exhibit root- $n$  convergence rate and circumvent the curse of dimensionality. The weak convergence of the process is obtained conditionally on a random direction, whilst it is proved the almost surely equivalence between the testing for significance expressed on the original and on the projected functional covariate. The computation of the test in practise involves the calibration by wild bootstrap resampling and the combination of several  $p$ -values arising from different projections by means of the false discovery rate method. The finite sample properties of the test are illustrated in a simulation study for a variety of linear models, underlying processes and alternatives. The software provided implements the tests and allows the replication of simulations and data applications.

**Keywords:** Empirical process; Functional data; Functional linear model; Functional principal components; Goodness-of-fit; Random projections.

## 1 Introduction

Since Karl Pearson introduced the term “goodness-of-fit” at the beginning of the twentieth century there has been an enormous amount of papers on this topic. First concentrated in fitting a model for one distribution function, and later, especially after the papers of Bickel and Rosenblatt (1973) and Durbin (1973), in more general models related with the regression function. The literature is vast, and we refer to González-Manteiga and Crujeiras (2013) for an updated review of the topic.

The ideas of goodness-of-fit for density and distribution have been naturally extended in the nineties of the last century to regression models. Considering, as a reference, a regression model with random design  $Y = m(X) + \varepsilon$  the goal is to test

$$H_0 : m \in \mathcal{M}_\Theta = \{m_\theta : \theta \in \Theta \subset \mathbb{R}^q\} \quad \text{vs.} \quad H_1 : m \notin \mathcal{M}_\Theta$$

in an omnibus way from a sample  $\{(X_i, Y_i)\}_{i=1}^n$  of  $(X, Y)$ . Here  $m(x) = \mathbb{E}[Y|X = x]$  is the regression function of  $Y$  over  $X$  and  $\varepsilon$  is a random error centred and such that  $\mathbb{E}[\varepsilon|\mathbf{X}] = 0$ .

Following the ideas on smoothing for testing about the density function (Bickel and Rosenblatt, 1973), the usual pilot estimator for  $m$  was a nonparametric one, for example, the Nadaraya-Watson estimator (Nadaraya (1964), Watson (1964)):  $\hat{m}_h(x) = \sum_{i=1}^n W_{ni}(x)Y_i$ , with  $W_{ni}(x) =$

<sup>1</sup>Department of Mathematics, Statistics and Computer Science, University of Cantabria (Spain).

<sup>2</sup>Department of Statistics, Carlos III University of Madrid (Spain).

<sup>3</sup>Department of Statistics, Mathematical Analysis and Operations Research, University of Santiago de Compostela (Spain).

<sup>4</sup>Corresponding author. e-mail: edgarcia@est-econ.uc3m.es.

$K((x - X_i)/h) / \sum_{j=1}^n K((x - X_j)/h)$ , where  $K$  is a kernel function and  $h$  is a bandwidth parameter. Other possible weights such as the ones from local linear estimation,  $k$ -nearest neighbours or splines were also used in different studies. Using these kind of pilot estimators, statistical tests were given by  $T_n = d(\hat{m}, m_{\hat{\theta}})$ , being  $d$  some functional distance and  $\hat{\theta}$  an estimator of  $\theta$  such that  $\sqrt{n}(\hat{\theta} - \theta) = \mathcal{O}_{\mathbb{P}}(1)$  under  $H_0$ . In an alternative way, following the paper by Durbin (1973) for testing about the distribution, the pilot estimator in the regression case was given by  $I_n(x) = n^{-1} \sum_{i=1}^n Y_i \mathbb{1}_{\{X_i \leq x\}}$ , the empirical estimation of the integrated regression function  $I(x) = \mathbb{E}[Y \mathbb{1}_{\{X \leq x\}}]$ . Härdle and Mammen (1993) using  $\hat{m}_h$  and Stute (1997) using  $I_n$  are key references for these two approximations in the literature, which were only the beginning of more than two hundred papers published in the last two decades (González-Manteiga and Crujeiras, 2013).

More recently, it has been of interest testing about a possible structure in a regression setting where we have functional covariates:

$$Y = m(\mathbf{X}) + \varepsilon, \quad (1)$$

being now  $\mathbf{X}$  a random element in a functional space, for example in the Hilbert space  $\mathcal{H} = L^2[0, 1]$ , and  $Y$  a scalar response. This is the context of “Functional Data Analysis”, which has received an increasing attention in the last decade (see for example Ramsay and Silverman (2005), Ferraty and Vieu (2006) and Horváth and Kokoszka (2012)), specially motivated by the practical needs of analysing data generated from high-resolution measuring devices.

A very simple null hypothesis  $H_0$  considered in the literature for the model (1) is  $H_0 : m(\mathbf{X}) = c$ , where  $c \in \mathbb{R}$  is a fixed constant: the testing of significance of the covariate  $\mathbf{X}$  over  $Y$ . Following some of the ideas from Ferraty and Vieu (2006) on considering pseudometrics for performing smoothing with functional data, the test by Härdle and Mammen (1993) was adapted by Delsol et al. (2011a):

$$T_n = \int (\hat{m}_h(\mathbf{x}) - \bar{Y})^2 \omega(\mathbf{x}) dP_{\mathbf{X}}(\mathbf{x}),$$

$$\hat{m}_h(\mathbf{x}) = \sum_{i=1}^n \left[ K\left(\frac{d(\mathbf{x}, \mathbf{X}_i)}{h}\right) Y_i / \sum_{j=1}^n K\left(\frac{d(\mathbf{X}_i, \mathbf{X}_j)}{h}\right) \right],$$

being  $d$  a functional pseudometric,  $K$  a kernel function adapted to this situation,  $h$  a bandwidth parameter,  $w$  a weight function and  $P_{\mathbf{X}}$  the probability measure induced by  $\mathbf{X}$  over the functional space of the covariate. The testing about  $H_0$  has been also considered by Cardot et al. (2003) or, more recently, Hilgert et al. (2013), not in an omnibus way but inside a Functional Linear Model (FLM):  $m(\mathbf{X}) = \langle \mathbf{X}, \boldsymbol{\rho} \rangle$ , where  $\langle \cdot, \cdot \rangle$  represents the inner product in  $\mathcal{H}$  and  $\boldsymbol{\rho} \in \mathcal{H}$  is the FLM parameter. For both approximations, omnibus or not, there has been also some recent papers considering the case of functional response; see for example, Cardot et al. (2007), Chiou and Müller (2007), Kokoszka et al. (2008) and Bücher et al. (2011).

The generalization of the previous functional hypothesis to the general case

$$H_0 : m \in \mathcal{M}_{\mathcal{P}} = \{m_{\boldsymbol{\rho}} : \boldsymbol{\rho} \in \mathcal{P}\} \quad \text{vs.} \quad H_1 : m \notin \mathcal{M}_{\mathcal{P}}, \quad (2)$$

where  $\mathcal{P}$  can be either of finite or infinite dimension, was the focus of very few papers, specially in the context of omnibus goodness-of-fit tests. In Delsol et al. (2011b) a discussion is given, without theoretical results, for the extension of the checking of a more complex null hypothesis such as a FLM. Only one paper is known for us where the FLM hypothesis is analysed with theoretical results. In Patilea et al. (2012), motivated by the smoothing test statistic considered by Zheng (1996) for finite dimensional covariates, one test based on

$$T_{n,h} = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} (Y_i - \hat{m}_{H_0}(\mathbf{X}_i)) (Y_j - \hat{m}_{H_0}(\mathbf{X}_j)) \frac{1}{h} K\left(\frac{F_{n,\mathbf{h}}(\langle \mathbf{X}_i, \mathbf{h} \rangle) - F_{n,\mathbf{h}}(\langle \mathbf{X}_j, \mathbf{h} \rangle)}{h}\right),$$

is employed for checking the null hypothesis of linearity with  $\hat{m}_{H_0}(\mathbf{X}) = \langle \mathbf{X}, \hat{\boldsymbol{\rho}} \rangle$ , with  $\hat{\boldsymbol{\rho}}$  a suitable estimator of  $\boldsymbol{\rho}$  and  $F_{n,\mathbf{h}}$  the empirical distribution function of  $\{\langle \mathbf{X}_i, \mathbf{h} \rangle\}_{i=1}^n$ . In the same spirit, Lavergne and Patilea (2008) gave a test for the finite dimensional context and Patilea et al. (2016) for functional response. From a different perspective, and motivated by the test of Escanciano (2006) for finite dimensional predictors, in García-Portugués et al. (2014) a test was constructed from the marked empirical process  $I_{n,\mathbf{h}}(x) = \frac{1}{n} \sum_{i=1}^n Y_i \mathbb{1}_{\{\langle \mathbf{X}_i, \mathbf{h} \rangle \leq x\}}$ , with  $x \in \mathbb{R}$  and  $\mathbf{h} \in \mathcal{H}$ . This approach circumvents the technical difficulties that a marked empirical process indexed by  $\mathbf{x} \in \mathcal{H}$ , a possible functional extension of the process in Stute (1997), would represent.

In this paper we consider marked empirical processes indexed by random projections of the functional covariate. The motivation stems from the almost surely (a.s.) characterization of the null hypothesis (2) via a *projected hypothesis* that arises from the conditional expectation on the projected functional covariate. This allows, conditionally on a randomly chosen  $\mathbf{h}$ , the study of the weak convergence of the process  $I_{n,\mathbf{h}}(x)$  for hypothesis testing of infinite dimension. As a by-product, we obtain root- $n$  goodness-of-fit tests that evade the curse of dimensionality and, contrary to smoothing-based tests, do not rely on a tuning parameter. Particularly, we focus on the testing of the aforementioned hypothesis of functional linearity where, contrary to the finite dimensional situation, the functional estimator has a non-trivial effect on the limiting process and requires a careful regularization. The test statistics are built by a continuous functional (Kolmogorov-Smirnov or Cramér-von Mises) over the empirical process and are effectively calibrated by a wild bootstrap on the residuals. To account for a higher power and less influence from  $\mathbf{h}$ , we consider a number  $K$  (not to confuse with a kernel function) of different random projections and merge the resulting  $p$ -values into a final  $p$ -value by means of the False Discovery Rate (FDR) of Benjamini and Yekutieli (2001). The empirical analysis reports a competitive performance of the test in practice, with a low impact of the choice of  $K$  above a certain bound, and an expedient computational complexity of  $\mathcal{O}(n)$  that yields notable speed improvements over García-Portugués et al. (2014).

The rest of the paper is organized as follows. The characterization of the null hypothesis through the projected predictor is addressed in Section 2, together with an application for the testing of the null hypothesis  $H_0 : m = m_0$  (Subsection 2.1). Section 3 is devoted to testing the composite hypothesis  $H_0 : m \in \{\langle \cdot, \boldsymbol{\rho} \rangle : \boldsymbol{\rho} \in \mathcal{H}\}$ . To that aim, the regularized estimator for  $\boldsymbol{\rho}$  of Cardot et al. (2007),  $\hat{\boldsymbol{\rho}}$ , is reviewed in Subsection 3.1. The pointwise asymptotic distribution of the projected process is studied in Subsection 3.2, whereas Subsection 3.3 gives its weak convergence. Section 4 describes the implementation of the test and other practicalities. Section 5 illustrates the finite sample properties of the test by a simulation study and with some real data applications. Some final comments and possible extensions are given in Section 6. Appendix A presents the main proofs, whereas the supplementary material contains the auxiliary lemmas and further results for the simulation study.

Some general setting and notation are introduced now. The random variable (r.v.)  $\mathbf{X}$  belongs to a separable Hilbert space  $\mathcal{H}$  endowed with the inner product  $\langle \cdot, \cdot \rangle$  and associated norm  $\|\cdot\|$ . The space  $\mathcal{H}$  is a general real Hilbert space but for simplicity can be regarded as  $\mathcal{H} = L^2[0, 1]$ .  $Y$  and  $\mathbf{X}$  are assumed to be centred r.v.'s, and  $\varepsilon$  is a centred, independent from  $\mathbf{X}$ , r.v. with variance  $\sigma_\varepsilon^2$ . The independence between  $\varepsilon$  and  $\mathbf{X}$  is a technical assumption required for proving Lemmas A.4 and A.5, while for the rest of the paper it suffices with  $\mathbb{E}[\varepsilon|\mathbf{X}] = 0$ . Given the  $\mathcal{H}$ -valued r.v.  $\mathbf{X}$  and  $\mathbf{h} \in \mathcal{H}$ , we denote by  $\mathbf{X}^{\mathbf{h}} = \langle \mathbf{X}, \mathbf{h} \rangle$  to the projected  $\mathbf{X}$  in the direction  $\mathbf{h}$ . Bold letters are used for vectors either in  $\mathcal{H}$  (mainly) or in  $\mathbb{R}^p$ , and its kind is clearly determined by the context. Capital letters represent r.v.'s defined on the same probability space  $(\Omega, \sigma, \nu)$ . Weak convergence is denoted by  $\xrightarrow{\mathcal{L}}$  and  $D(\mathbb{R})$  represents the Skorohod's space of càdlàg functions defined on  $\mathbb{R}$ . Finally, we shall implicitly assume that the null hypotheses stated hold a.s.

## 2 Hypothesis projection

The pillar of the goodness-of-fit tests we present is the a.s. characterization of the null hypothesis (2), re-expressed as  $H_0 : \mathbb{E}[Y - m_{\boldsymbol{\rho}}(\mathbf{X})|\mathbf{X}] = 0$  for some  $\boldsymbol{\rho} \in \mathcal{P}$ , by means of the associated *projected hypothesis on  $\mathbf{h} \in \mathcal{H}$* , defined as  $H_0^{\mathbf{h}} : \mathbb{E}[Y - m_{\boldsymbol{\rho}}(\mathbf{X})|\mathbf{X}^{\mathbf{h}}] = 0$ . We identify  $Y - m_{\boldsymbol{\rho}}(\mathbf{X})$  by  $Y$  for the sake of simplicity in notation. We give in this section two necessary and sufficient conditions based on the projections of  $\mathbf{X}$  to that  $\mathbb{E}[Y|\mathbf{X}] = 0$  holds a.s.

The first condition only requires the integrability of  $Y$ , but the condition needs to be satisfied for every direction  $\mathbf{h}$ .

**Proposition 2.1.** *Assume that  $\mathbb{E}[|Y|] < \infty$ . Then,*

$$\mathbb{E}[Y|\mathbf{X}] = 0 \text{ a.s.} \iff \mathbb{E}[Y|\mathbf{X}^{\mathbf{h}}] = 0 \text{ a.s. for every } \mathbf{h} \in \mathcal{H}.$$

The second condition, more adequate for application, *somehow* generalizes Proposition 2.1, as it only needs to be satisfied for a randomly chosen  $\mathbf{h}$ . In exchange, it holds only under some additional conditions on the moments of  $\mathbf{X}$  and  $Y$ . Before stating it we need some preliminary results, being the first one included here for the sake of completeness.

**Lemma 2.2** (Theorem 4.1 in Cuesta-Albertos et al. (2007)). *Let  $\mu$  be a non-degenerate Gaussian measure on  $\mathcal{H}$ , let  $\mathbf{X}_1, \mathbf{X}_2$  be two  $\mathcal{H}$ -valued r.v.'s and denote by  $\mathbf{X}_1 \sim \mathbf{X}_2$  if they are identically distributed. Assume that:*

- (a)  $m_k := \int \|\mathbf{X}_1\|^k d\nu < \infty$ , for all  $k \geq 1$ , and  $\sum_{k=1}^{\infty} m_k^{-1/k} = \infty$ .
- (b) The set  $\{\mathbf{h} \in \mathcal{H} : \mathbf{X}_1^{\mathbf{h}} \sim \mathbf{X}_2^{\mathbf{h}}\}$  is of positive  $\mu$ -measure.

Then  $\mathbf{X}_1 \sim \mathbf{X}_2$ .

**Remark 2.2.1.** It is not strictly needed that  $\mu$  be a Gaussian distribution in Lemma 2.2 and this can be replaced by assuming a certain smoothness condition on  $\mu$  (see, for instance, Theorem 2.5 and Example 2.6 in Cuesta-Albertos et al. (2007)).

**Remark 2.2.2.** Assumption (a) in Lemma 2.2 is not of technical nature. According to Theorem 3.6 in Cuesta-Albertos et al. (2007), it becomes apparent that a similar condition is required. This assumption is satisfied if the tails of  $P_{\mathbf{X}_1}$  are light enough or if  $\mathbf{X}_1$  has a finite moment generating function in a neighbourhood of zero.

**Lemma 2.3.** *If  $\mathbb{E}[Y^2] < \infty$  and  $\mathbf{X}$  satisfies (a) in Lemma 2.2, then  $l_k := \mathbb{E}[\|\mathbf{X}\|^k |Y|] < \infty$ , for all  $k \geq 1$ , and  $\sum_{k=1}^{\infty} l_k^{-1/k} = \infty$ .*

The second condition and most important result in this section is given as follows.

**Theorem 2.4.** *Let  $\mu$  be a non-degenerate Gaussian measure on  $\mathcal{H}$ . Assume that  $\mathbf{X}$  satisfies (a) in Lemma 2.2 and that  $\mathbb{E}[Y^2] < \infty$ . Then,*

$$\mathbb{E}[Y|\mathbf{X}] = 0 \text{ a.s.} \iff \mathcal{H}_0 := \{\mathbf{h} \in \mathcal{H} : \mathbb{E}[Y|\mathbf{X}^{\mathbf{h}}] = 0 \text{ a.s.}\} \text{ has positive } \mu\text{-measure.}$$

**Corollary 2.5.** *Under the assumptions of the previous theorem,*

$$\mathbb{E}[Y|\mathbf{X}] = 0 \text{ a.s.} \iff \mu(\mathcal{H}_0) = 1.$$

According to this corollary, it happens that if we are interested in testing the simple null hypothesis  $H_0 : \mathbb{E}[Y|\mathbf{X}] = 0$  we can do it as follows: i) select at random with  $\mu$  a direction  $\mathbf{h} \in \mathcal{H}$ ; ii) conditionally on  $\mathbf{h}$ , test the projected null hypothesis  $H_0^{\mathbf{h}} : \mathbb{E}[Y|\mathbf{X}^{\mathbf{h}}] = 0$ . The rationale is simple yet powerful: if  $H_0$  holds, then  $H_0^{\mathbf{h}}$  also holds; if  $H_0$  fails, then  $H_0^{\mathbf{h}}$  also fails  $\mu$ -a.s. In this case, with probability one we have chosen a direction  $\mathbf{h}$  for which  $H_0^{\mathbf{h}}$  fails. Of course, the main advantage to test  $H_0^{\mathbf{h}}$  over testing  $H_0$  directly is that in  $H_0^{\mathbf{h}}$  the conditioning r.v. is real, which simplifies the problem substantially.

## 2.1 Testing a simple null hypothesis

An immediate application of Corollary 2.5 is the testing of the simple null hypothesis  $H_0 : m = m_0$  via the empirical process of Stute (1997). Recall that other testing alternatives can be considered on the projected covariate due to the  $\mu$ -a.s. characterization. We refer to González-Manteiga and Crujeiras (2013) for a review on alternatives.

For a random sample  $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$  from  $(\mathbf{X}, Y)$ , we can consider the empirical process of the regression conditioned on the direction  $\mathbf{h}$ ,

$$R_{n,\mathbf{h}}(x) := n^{1/2} I_{n,\mathbf{h}}(x) = n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} Y_i, \quad x \in \mathbb{R}$$

and then the following result is trivially satisfied from Theorem 1.1 in Stute (1997).

**Corollary 2.6.** *Under  $H_0^{\mathbf{h}}$  and  $\mathbb{E}[Y^2] < \infty$ ,  $R_{n,\mathbf{h}} \xrightarrow{\mathcal{L}} \mathcal{G}_1$  in  $D(\mathbb{R})$ , being  $\mathcal{G}_1$  a Gaussian process with zero mean and covariance function  $K_1(s, t) := \int_{-\infty}^{s \wedge t} \text{Var}[Y | \mathbf{X}^{\mathbf{h}} = u] dF_{\mathbf{h}}(u)$ , where  $F_{\mathbf{h}}$  is the distribution function of  $\mathbf{X}^{\mathbf{h}}$ .*

Different statistics for the testing of  $H_0^{\mathbf{h}}$  can be built from continuous functionals on  $R_{n,\mathbf{h}}(x)$ . We shall cover this with more detail in Section 3.

**Example 2.7.** Consider the FLM  $Y = \langle \mathbf{X}, \boldsymbol{\rho} \rangle + \varepsilon$  in  $\mathcal{H} = L^2[0, 1]$ , with  $\mathbf{X}$  a Gaussian process with associated Karhunen-Loève expansion (4) and  $\varepsilon$  independent from  $\mathbf{X}$ . Then  $\mathbf{X}^{\mathbf{h}}$  and  $\mathbf{X}^{\boldsymbol{\rho}}$  are centred Gaussians with variances  $\sigma_{\mathbf{h}}^2$  and  $\sigma_{\boldsymbol{\rho}}^2$ , respectively, and  $\text{Cov}[\mathbf{X}^{\mathbf{h}}, \mathbf{X}^{\boldsymbol{\rho}}] = \sum_{j=1}^{\infty} h_j \rho_j \lambda_j$ , with  $h_j = \langle \mathbf{h}, \mathbf{e}_j \rangle$ ,  $\rho_j = \langle \boldsymbol{\rho}, \mathbf{e}_j \rangle$ . Hence,

$$\begin{aligned} K^1(s, t) &= \int_{-\infty}^{s \wedge t} \left( \frac{\sigma_{\boldsymbol{\rho}}^2 \sigma_{\mathbf{h}}^2 - \left( \sum_{j=1}^{\infty} h_j \rho_j \lambda_j \right)^2}{\sigma_{\mathbf{h}}^2} + \sigma_{\varepsilon}^2 \right) \phi(u / \sigma_{\mathbf{h}}) / \sigma_{\mathbf{h}} du \\ &= \left( \frac{\sigma_{\boldsymbol{\rho}}^2 \sigma_{\mathbf{h}}^2 - \left( \sum_{j=1}^{\infty} h_j \rho_j \lambda_j \right)^2}{\sigma_{\mathbf{h}}^2} + \sigma_{\varepsilon}^2 \right) \Phi((s \wedge t) / \sigma_{\mathbf{h}}), \end{aligned}$$

where  $\phi$  and  $\Phi$  are the density and distribution functions of a  $\mathcal{N}(0, 1)$ , respectively.

## 3 Testing the functional linear model

We focus now in testing the composite null hypothesis, expressed as

$$H_0 : m(\mathbf{X}) = \langle \mathbf{X}, \boldsymbol{\rho} \rangle = \mathbf{X}^{\boldsymbol{\rho}}, \text{ for some } \boldsymbol{\rho} \in \mathcal{H}.$$

According to Corollary 2.5, it happens that testing  $H_0$  is  $\mu$ -a.s. equivalent to test

$$H_0^{\mathbf{h}} : \mathbb{E}[(Y - \mathbf{X}^{\boldsymbol{\rho}}) | \mathbf{X}^{\mathbf{h}}] = 0, \text{ for some } \boldsymbol{\rho} \in \mathcal{H},$$

where  $\mathbf{h}$  is sampled from a non-degenerate Gaussian law  $\mu$ . Again, we construct the associated empirical regression process indexed by the projected covariate following Stute (1997). Therefore, given an estimate  $\hat{\boldsymbol{\rho}}$  of  $\boldsymbol{\rho}$  under  $H_0$ , we have

$$T_{n,\mathbf{h}}(x) := a_n \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} (Y_i - \mathbf{X}_i^{\hat{\boldsymbol{\rho}}}) = a_n (T_{n,\mathbf{h}}^1(x) + T_{n,\mathbf{h}}^2(x) + T_{n,\mathbf{h}}^3(x)), \quad (3)$$

where  $a_n \rightarrow 0$  is a normalizing positive sequence to be determined later and

$$\begin{aligned} T_{n,\mathbf{h}}^1(x) &:= \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} (Y_i - \mathbf{X}_i^{\boldsymbol{\rho}}), \\ T_{n,\mathbf{h}}^2(x) &:= \sum_{i=1}^n \left\langle \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \mathbf{X}_i - \mathbb{E} \left[ \mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \leq x\}} \mathbf{X} \right], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}} \right\rangle, \\ T_{n,\mathbf{h}}^3(x) &:= n \left\langle \mathbb{E} \left[ \mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \leq x\}} \mathbf{X} \right], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}} \right\rangle. \end{aligned}$$

The selection of the right estimator  $\hat{\boldsymbol{\rho}}$  has a crucial role in the weak convergence of  $T_{n,\mathbf{h}}^3$ , which poses a substantially complexer proof than for the simple hypothesis. We consider the regularized estimate proposed in Sections 2 and 3 of Cardot et al. (2007) (denoted by CMS in the sequel), whose construction is sketched here for the sake of exposition of our results.

### 3.1 Construction of the estimator of $\boldsymbol{\rho}$

Consider the so-called Karhunen-Loève expansion of  $\mathbf{X}$ :

$$\mathbf{X} = \sum_{j=1}^{\infty} \lambda_j^{1/2} \xi_j \mathbf{e}_j, \quad (4)$$

where  $\{\mathbf{e}_j\}_{j=1}^{\infty}$  is a sequence of orthonormal eigenfunctions associated to the covariance operator of  $\mathbf{X}$ ,  $\Gamma \mathbf{z} := \mathbb{E}[(\mathbf{X} \otimes \mathbf{X})(\mathbf{z})] = \mathbb{E}[\langle \mathbf{z}, \mathbf{X} \rangle \mathbf{X}]$ ,  $\mathbf{z} \in \mathcal{H}$  and the  $\xi_j$ 's are centred real r.v.'s (because  $\mathbf{X}$  is centred) such that  $\mathbb{E}[\xi_j \xi_{j'}] = \delta_{j,j'}$ , where  $\delta_{j,j'}$  is the Kronecker's delta. We assume that the multiplicity of each eigenvalue is one, so  $\lambda_1 > \lambda_2 > \dots > 0$ .

The functional coefficient  $\boldsymbol{\rho}$  is determined by the equation  $\Delta = \Gamma \boldsymbol{\rho}$ , with  $\Delta$  the cross-covariance operator of  $\mathbf{X}$  and  $Y$ ,  $\Delta \mathbf{z} := \mathbb{E}[(\mathbf{X} \otimes Y)(\mathbf{z})] = \mathbb{E}[\langle \mathbf{z}, \mathbf{X} \rangle Y]$ ,  $\mathbf{z} \in \mathcal{H}$ . To ensure the existence and uniqueness of a solution to  $\Delta = \Gamma \boldsymbol{\rho}$  we require the next basic assumptions:

**A1.**  $\mathbf{X}$  and  $Y$  satisfy  $\sum_{j=1}^{\infty} \frac{1}{\lambda_j^2} \langle \mathbb{E}[\mathbf{X}Y], \mathbf{e}_j \rangle^2 < \infty$ .

**A2.**  $\text{Ker}(\Gamma) = \{\mathbf{0}\}$ .

The estimation of  $\boldsymbol{\rho}$  requires the inversion of  $\Gamma_n := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \otimes \mathbf{X}_i$ , but since  $\Gamma_n$  is a.s. a finite rank operator, its inverse does not exist. CMS proposed a regularization yielding a family of continuous estimators for  $\Gamma^{-1}$ . We employ the one from Example 1 in CMS, which provides an empirical finite rank inverse of  $\Gamma$  denoted  $\Gamma_n^{\dagger}$  (we use  $\Gamma^{\dagger}$  for the population version). Consider a sequence of thresholds  $c_n \in (0, \lambda_1)$ ,  $n \in \mathbb{N}$ , with  $c_n \rightarrow 0$ . Then: *i*) compute the Functional Principal Components (FPC) of  $\Gamma_n$ , *i.e.*, calculate its eigenvalues  $\{\hat{\lambda}_j\}$  and eigenfunctions  $\{\hat{\mathbf{e}}_j\}$ ; *ii*) define the sequences  $\{\delta_j\}$ , with  $\delta_1 := \lambda_1 - \lambda_2$  and  $\delta_j := \min(\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j)$  for  $j > 1$ , and set

$$k_n := \sup\{j \in \mathbb{N} : \lambda_j + \delta_j/2 \geq c_n\};$$

*iii*) compute  $\Gamma_n^{\dagger}$  (respectively  $\Gamma^{\dagger}$ ) as the finite rank operator with the same eigenfunctions as  $\Gamma_n$  (resp.  $\Gamma$ ) and associated eigenvalues equal to  $\hat{\lambda}_j^{-1}$  (resp.  $\lambda_j^{-1}$ ) if  $j \leq k_n$  and 0 otherwise. The regularized estimator of  $\boldsymbol{\rho}$  is

$$\hat{\boldsymbol{\rho}} := \Gamma_n^{\dagger} \Delta_n = \frac{1}{n} \sum_{j=1}^{k_n} \sum_{i=1}^n \frac{\langle \mathbf{X}_i \otimes Y_i, \hat{\mathbf{e}}_j \rangle}{\hat{\lambda}_j} \hat{\mathbf{e}}_j. \quad (5)$$

Note that (5) is not readily computable in practise, since  $\{\lambda_j\}$  is usually unknown (and hence  $k_n$ ). As in CMS, we consider the (random) finite rank

$$d_n := \sup\{j \in \mathbb{N} : \hat{\lambda}_j \geq c_n\}$$

as a replacement in practise for the deterministic  $k_n$ . As seen in Lemma A.2,  $\nu[k_n = d_n] \rightarrow 1$ , hence the estimator (5) has the same asymptotic behaviour with either  $k_n$  or  $d_n$ . Therefore, we consider  $k_n$  in (5) due to the convenient probabilistic tractability.

The following assumptions allow to obtain meaningful asymptotic convergences:

**A3.**  $\mathbb{E} [\|\mathbf{X}\|^2] < \infty$ .

**A4.**  $\sum_{l=1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_l \rangle| < \infty$ .

**A5.** For  $j$  large,  $\lambda_j = \lambda(j)$  with  $\lambda(\cdot)$  a convex positive function.

**A6.**  $\frac{\lambda_n n^4}{\log n} = \mathcal{O}(1)$ .

**A7.**  $\inf \left\{ |\langle \boldsymbol{\rho}, \mathbf{e}_{k_n} \rangle|, \frac{\lambda_{k_n}}{\sqrt{k_n \log k_n}} \right\} = \mathcal{O}(n^{-1/2})$ .

**A8.**  $\sup_j \left\{ \max \left( \mathbb{E}[\xi_j^4], \mathbb{E}[|\xi_j|^5] \right) \right\} \leq M < \infty$ , for  $M \geq 1$ .

**A9.**  $c_n = \mathcal{O}(n^{-1/2})$ .

A brief summary of these assumptions is given as follows. **A3** is standard to obtain asymptotic distributions, allows decomposition (4) and implies  $\mathbb{E}[Y^2] < \infty$ , required in Theorem 1.1 of Stute (1997). **A4** and **A5** are A.1 and A.2 in CMS. **A6** is very similar to one assumption in the second part of Theorem 2 in CMS. **A7** is the minimum requirement to control  $\langle \mathbf{X}, \mathbf{L}_n \rangle$  when Lemma 7 in CMS is used to prove Lemma A.7. **A8** is a reinforcement of A.3 in CMS, where only fourth order moments are used. The reason is because we handle inner products of  $\hat{\boldsymbol{\rho}}$  times a non independent r.v., while in CMS the r.v. is not used to estimate  $\boldsymbol{\rho}$ . **A9** is useful, mainly (but also see the final part of Lemma A.6) to control the behaviour of  $k_n$ . We show this fact in Proposition A.1, with a conclusion very close to the assumption (8) in CMS and coinciding with one of the conditions of their Theorem 3 if  $\lim_n t_{n, \mathbf{E}_{x, \mathbf{h}}} < \infty$  (the term  $t_{n, \mathbf{E}_{x, \mathbf{h}}}$  is defined in (6)). Finally, we point out that in CMS the assumptions aim to control the behaviour of  $k_n$  while here we have targeted to control the threshold  $c_n$ , as this can be modified by the statistician.

### 3.2 Pointwise asymptotic distribution of $T_{n, \mathbf{h}}$

Corollary 2.6 gives the weak convergence of  $n^{-1/2} T_{n, \mathbf{h}}^1$ . We analyse now the pointwise behaviour of  $T_{n, \mathbf{h}}^2(x)$  and  $T_{n, \mathbf{h}}^3(x)$  for a fixed  $x \in \mathbb{R}$ . We will show that  $T_{n, \mathbf{h}}^2(x) = o_{\mathbb{P}}(n^{1/2})$  and that the rate of  $T_{n, \mathbf{h}}^3(x)$  depends on the key normalizing sequence  $\{t_{n, \mathbf{E}_{x, \mathbf{h}}}\}$ , where

$$t_{n, \mathbf{x}} := \sqrt{\sum_{j=1}^{k_n} \frac{\langle \mathbf{x}, \mathbf{e}_j \rangle^2}{\lambda_j}} \quad \text{and} \quad \mathbf{E}_{x, \mathbf{h}} := \mathbb{E} \left[ \mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \leq x\}} \mathbf{X} \right]. \quad (6)$$

**Theorem 3.1.** *Under  $H_0^{\mathbf{h}}$  and **A1–A9**, for a fixed  $x \in \mathbb{R}$ , it happens that:*

- (a)  $n^{-1/2} t_{n, \mathbf{E}_{x, \mathbf{h}}}^{-1} T_{n, \mathbf{h}}^3(x) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{\varepsilon}^2)$ .
- (b) If  $\lim_n t_{n, \mathbf{E}_{x, \mathbf{h}}} = \infty$ , then with  $a_n = n^{-1/2} t_{n, \mathbf{E}_{x, \mathbf{h}}}^{-1}$  in (3) the asymptotic distribution of  $T_{n, \mathbf{h}}(x)$  is the one of  $n^{-1/2} t_{n, \mathbf{E}_{x, \mathbf{h}}}^{-1} T_{n, \mathbf{h}}^3(x)$ .
- (c) If  $\lim_n t_{n, \mathbf{E}_{x, \mathbf{h}}} < \infty$ , then with  $a_n = n^{-1/2}$  in (3) the asymptotic distribution of  $T_{n, \mathbf{h}}(x)$  is the one of  $n^{-1/2} (T_{n, \mathbf{h}}^1(x) + T_{n, \mathbf{h}}^3(x))$ .

The behaviour of the sequence  $\{t_{n,\mathbf{E}_{x,\mathbf{h}}}\}$ , indexed by  $n \in \mathbb{N}$  and with arbitrary  $\mathbf{h} \in \mathcal{H}$  and  $x \in \mathbb{R}$ , is crucial for the convergence of  $T_{n,\mathbf{h}}$ . Since  $\{t_{n,\mathbf{E}_{x,\mathbf{h}}}\}$  is non-decreasing, it has always a limit (finite or infinite). Its asymptotic behaviour is described next.

**Proposition 3.2.** *The sequence  $\{t_{n,\mathbf{E}_{x,\mathbf{h}}}\}$  has asymptotic orders between  $\mathcal{O}(1)$  and  $\mathcal{O}(k_n^{1/2})$ . In addition, if  $\mathbf{X}$  is Gaussian and satisfies **A3**, then  $\sigma_{\mathbf{h}}^2 := \text{Var}[\mathbf{X}^{\mathbf{h}}] < \infty$  and  $\lim_n t_{n,\mathbf{E}_{x,\mathbf{h}}} = \phi(x/\sigma_{\mathbf{h}})$ .*

### 3.3 Weak convergence of $T_{n,\mathbf{h}}$ and of the test statistics

The result given in Theorem 3.1 holds for every  $x \in \mathbb{R}$ . For the case (c) of Theorem 3.1 (where the estimation of  $\boldsymbol{\rho}$  is not dominant) and under an additional assumption, the result can be generalized to functional weak convergence.

**Theorem 3.3.** *Under  $H_0^{\mathbf{h}}$ , **A1–A9** and (c) in Theorem 3.1, it happens that:*

- (a) *The finite dimensional distributions of  $T_{n,\mathbf{h}}$  converge to a multivariate Gaussian with covariance function  $K_2(s, t) := K_1(s, t) + C(s, t) + C(t, s) + V(s, t)$ , where*

$$C(s, t) := \int_{\{\mathbf{u}^{\mathbf{h}} \leq s\}} \text{Var}[Y|\mathbf{X} = \mathbf{u}] \langle \mathbf{E}_{t,\mathbf{h}}, \Gamma^{-1} \mathbf{u} \rangle dP_{\mathbf{X}}(\mathbf{u}),$$

$$V(s, t) := \int \text{Var}[Y|\mathbf{X} = \mathbf{u}] \langle \mathbf{E}_{s,\mathbf{h}}, \Gamma^{-1} \mathbf{u} \rangle \langle \mathbf{E}_{t,\mathbf{h}}, \Gamma^{-1} \mathbf{u} \rangle dP_{\mathbf{X}}(\mathbf{u}).$$

- (b) *If  $\mathbb{E}[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|^4] = \mathcal{O}(n^{-2})$ , then  $T_{n,\mathbf{h}} \xrightarrow{\mathcal{L}} \mathcal{G}_2$  in  $D(\mathbb{R})$ , being  $\mathcal{G}_2$  a Gaussian process with zero mean and covariance function  $K_2$ .*

**Remark 3.3.1.** *According to Theorem 1 in CMS, it is impossible for  $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}$  to converge to a non-degenerate random element in the topology of  $\mathcal{H}$ . In order to circumvent this issue, we make the assumption  $\mathbb{E}[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|^4] = \mathcal{O}(n^{-2})$  which implies  $\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\| = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ , thus a finite-dimensional parametric convergence rate for  $\hat{\boldsymbol{\rho}}$ . In practice this means that  $\boldsymbol{\rho}$  lives in a finite dimensional subspace of  $\mathcal{H}$ .*

The next result gives the convergence of the Kolmogorov-Smirnov (KS) and Cramér-von Mises (CvM) statistics for testing the FLM.

**Corollary 3.4.** *Under the assumptions in Theorem 3.3 and  $\mathbb{E}[\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|^4] = \mathcal{O}(n^{-2})$ , if  $\|T_{n,\mathbf{h}}\|_{\text{KS}} := \sup_{x \in \mathbb{R}} |T_{n,\mathbf{h}}(x)|$  and  $\|T_{n,\mathbf{h}}\|_{\text{CvM}} := \int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2 dF_{\mathbf{h}}(x)$ , then*

$$\|T_{n,\mathbf{h}}\|_{\text{KS}} \xrightarrow{\mathcal{L}} \|\mathcal{G}_2\|_{\text{KS}} \text{ and } \|T_{n,\mathbf{h}}\|_{\text{CvM}} \xrightarrow{\mathcal{L}} \int_{\mathbb{R}} \mathcal{G}_2(x)^2 dF_{\mathbf{h}}(x).$$

**Remark 3.4.1.** *An alternative to (b) and Corollary 3.4 is to consider a deterministic discretization of the statistics, for which the convergence in law is trivial from (a). For example, if  $\|T_{n,\mathbf{h}}\|_{\widehat{\text{KS}}} := \max_{k=1,\dots,G} |T_{n,\mathbf{h}}(x_k)|$  for a grid  $\{x_1, \dots, x_G\}$ , then  $\|T_{n,\mathbf{h}}\|_{\widehat{\text{KS}}} \xrightarrow{\mathcal{L}} \|\mathbf{Z}_2\|_{\widehat{\text{KS}}}$ , where  $\mathbf{Z}_2 \sim \mathcal{N}_G(\mathbf{0}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\Sigma}_{ij} = K_2(x_i, x_j)$ .*

## 4 Testing in practise

The major advantage to test  $H_0^{\mathbf{h}}$  over  $H_0$  is that in  $H_0^{\mathbf{h}}$  the conditioning r.v. is real. The potential drawbacks of this universal method are a possible loss of power and that the outcome of the test may vary for different projections. Both inconveniences can be alleviated by sampling several directions  $\mathbf{h}_1, \dots, \mathbf{h}_K$ , testing the projected hypotheses  $H_0^{\mathbf{h}_1}, \dots, H_0^{\mathbf{h}_K}$  and selecting an appropriate way to mix the resulting  $p$ -values. For example, by the FDR method proposed in Benjamini and Yekutieli (2001) (see Section 2.2.2 of Cuesta-Albertos and Febrero-Bande (2010)) it is possible to control the final rejection rate to be *at most*  $\alpha$  under  $H_0$ . The procedure is described in the following generic algorithm.



**Algorithm 4.1** (Testing procedure for  $H_0$ ). Let  $T_n$  denote a test for checking  $H_0^{\mathbf{h}}$  with  $\mathbf{h}$  chosen by a non-degenerate Gaussian measure  $\mu$  on  $\mathcal{H}$ .

- i) For  $i = 1, \dots, K$ , set by  $p_i$  the  $p$ -value of  $H_0^{\mathbf{h}_i}$  obtained with the test  $T_n$ .
- ii) Set the final  $p$ -value of  $H_0$  as  $\min_{i=1, \dots, K} \frac{K}{i} p_{(i)}$ , where  $p_{(1)} \leq \dots \leq p_{(K)}$ .

The calibration of the test statistic for  $H_0^{\mathbf{h}}$  is done by a wild bootstrap resampling. The next algorithm states the steps for testing the FLM. The particular case of the simple null hypothesis corresponds to  $\boldsymbol{\rho} = \mathbf{0}$ , so its calibration corresponds to setting  $\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\rho}}^* = \mathbf{0}$  in the algorithm.

**Algorithm 4.2** (Bootstrap calibration in FLM testing). Let  $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$  be a random sample from (1). To test  $H_0 : m \in \{\langle \cdot, \boldsymbol{\rho} \rangle : \boldsymbol{\rho} \in \mathcal{H}\}$  proceed as follows:

- i) Estimate  $\boldsymbol{\rho}$  by FPC for a given  $d_n$  and obtain  $\hat{\varepsilon}_i = Y_i - \langle \mathbf{X}_i, \hat{\boldsymbol{\rho}} \rangle$ .
- ii) Compute  $\|T_{n,\mathbf{h}}\|_N = \left\| n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i \right\|_N$  with  $N$  either KS or CvM.
- iii) Bootstrap resampling. For  $b = 1, \dots, B$ , do:
  - a) Draw binary i.i.d. r.v.'s  $V_1^*, \dots, V_n^*$  such that  $\mathbb{P}\{V^* = (1 - \sqrt{5})/2\} = (5 + \sqrt{5})/10$  and  $\mathbb{P}\{V^* = (1 + \sqrt{5})/2\} = (5 - \sqrt{5})/10$ .
  - b) Set  $Y_i^* := \langle \mathbf{X}_i, \hat{\boldsymbol{\rho}} \rangle + \varepsilon_i^*$  from the bootstrap residuals  $\varepsilon_i^* := V_i^* \hat{\varepsilon}_i$ .
  - c) Estimate  $\boldsymbol{\rho}^*$  from  $\{(\mathbf{X}_i, Y_i^*)\}_{i=1}^n$  by FPC using the same  $d_n$  of  $i$ .
  - d) Obtain the estimated bootstrap residuals  $\hat{\varepsilon}_i^* := Y_i^* - \langle \mathbf{X}_i, \hat{\boldsymbol{\rho}}^* \rangle$ .
  - e) Compute  $\|T_{n,\mathbf{h}}^{*b}\|_N := \left\| n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i^* \right\|_N$ .
- iv) Approximate the  $p$ -value by  $\frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\|T_{n,\mathbf{h}}^{*b}\|_N \leq \|T_{n,\mathbf{h}}\|_N\}}$ .

The choice of an adequate  $d_n$  for the estimation of  $\boldsymbol{\rho}$  can be done in a data-driven way, for example by the corrected Schwartz Information Criterion (McQuarrie, 1999), denoted by SICc. Besides, Steps c) and d) can be easily computed using the properties of the linear model, see Section 3.3 of García-Portugués et al. (2014).

The bootstrap process we are considering is given by (we consider  $a_n = n^{-1/2}$ ):

$$T_{n,\mathbf{h}}^*(x) := n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i^* = n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i V_i^* + n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \mathbf{X}_i^{\hat{\boldsymbol{\rho}} - \hat{\boldsymbol{\rho}}^*}$$

which is estimating the distribution of

$$T_{n,\mathbf{h}}(x) = n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \hat{\varepsilon}_i + n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \mathbf{X}_i^{\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}}.$$

The bootstrap consistency could be obtained as an adaptation of: Lemma A.1 of Stute et al. (1998) for the first term of  $T_{n,\mathbf{h}}^*$ ; Lemma A.2 of the same paper for the second term, using the decomposition of  $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}$  given in (11) of CMS.

The drawing of the random directions is clearly influential in the power of the test. For example, in the extreme case where the projections were orthogonal to the data, that is  $\mathbf{X}^{\mathbf{h}} = 0$ , then  $T_{n,\mathbf{h}}(x) = (n^{-1/2} \sum_{i=1}^n \hat{\varepsilon}_i) \mathbb{1}_{\{0 \leq x\}}$  and  $\|T_{n,\mathbf{h}}\|_N = \|T_{n,\mathbf{h}}^*\|_N = 0$  under  $H_0$ . Therefore, Algorithm 4.2 would fail to calibrate the level of the test and potentially yield spurious results due to numerical inaccuracies in  $\|T_{n,\mathbf{h}}^{*b}\|_N \leq \|T_{n,\mathbf{h}}\|_N$ . A data-driven compromise to avoid drawing projections in

subspaces *almost* orthogonal to the data is the following: *i*) compute the FPC of  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , *i.e.*, the eigenpairs  $\{(\hat{\lambda}_j, \hat{\mathbf{e}}_j)\}$ ; *ii*) choose  $j_n := \min \{k = 1, \dots, n-1 : (\sum_{j=1}^k \hat{\lambda}_j^2)/(\sum_{j=1}^{n-1} \hat{\lambda}_j^2) \geq r\}$  for a variance threshold  $r$ , e.g.  $r = 0.95$ ; *iii*) generate the data-driven Gaussian process  $\mathbf{h}_{j_n} := \sum_{j=1}^{j_n} \eta_j \hat{\mathbf{e}}_j$ , with  $\eta_j \sim \mathcal{N}(0, s_j^2)$  and  $s_j^2$  the sample variance of the scores in the  $j$ -th FPC. Formally, the Gaussian measure  $\mu$  associated to  $\mathbf{h}_{j_n}$  does not respect the assumptions in Theorem 2.4, since it is degenerate (but recall that  $\mu$  does not have to be independent from  $\mathbf{X}$ ). A non-degenerate Gaussian process can be obtained as  $\mathbf{h}_{j_n} + \mathcal{G}$ , with  $\mathcal{G}$  a Gaussian process tightly concentrated around zero, albeit employing  $\mathbf{h}_{j_n}$  or  $\mathbf{h}_{j_n} + \mathcal{G}$  has negligible effects in practise.

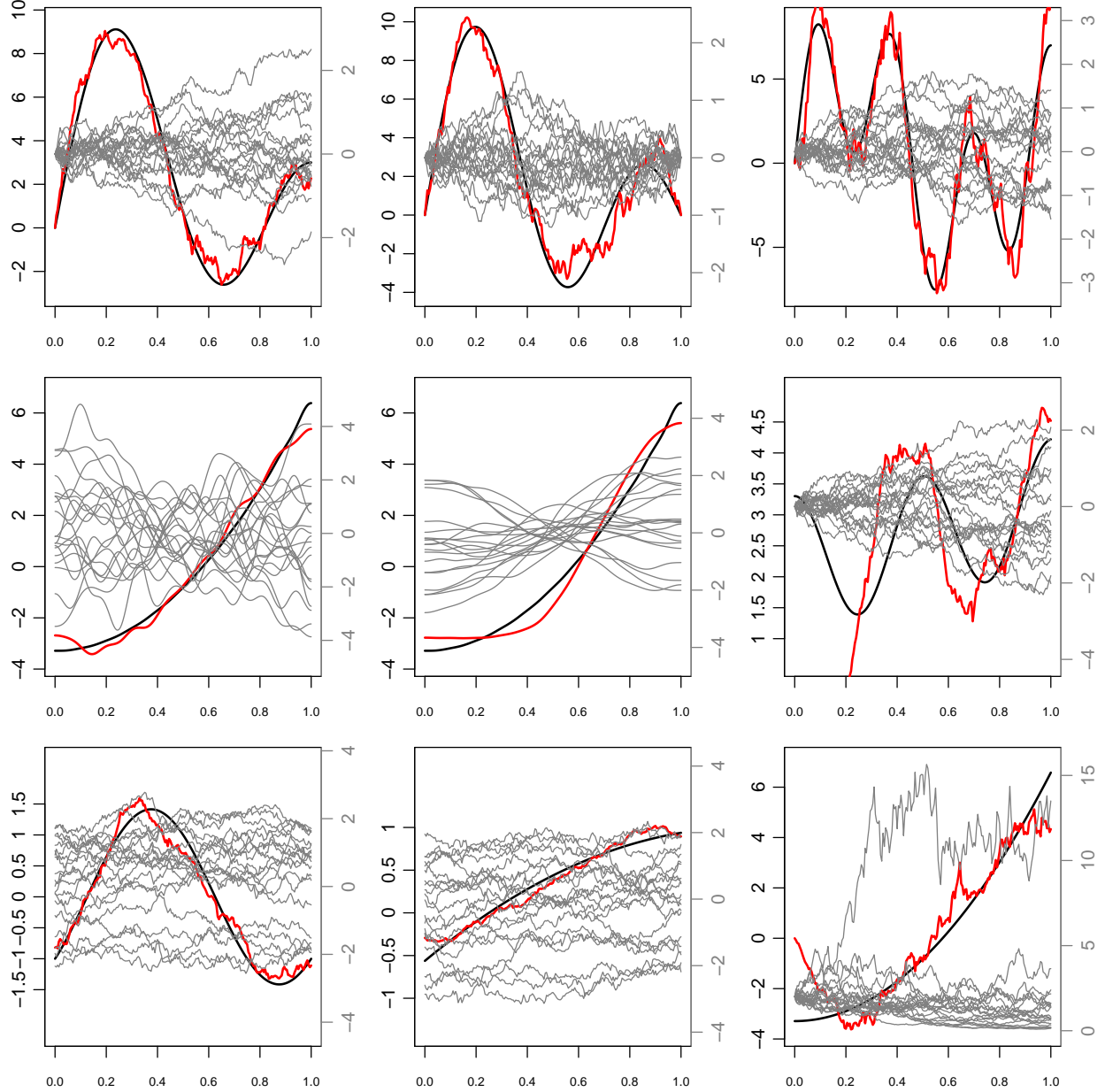


Figure 1: From left to right and up to down, functional coefficients  $\rho$  (black, right scale) and underlying processes (grey, left scale) for the nine different scenarios, labelled S1 to S9. Each graph contains a sample of 100 realizations of the functional covariate  $\mathbf{X}$  and  $\hat{\rho}$  (red) with  $d_n$  selected by SICc.

## 5 Simulation study and data application

We illustrate the finite sample performance of the CvM and KS goodness-of-fit tests implemented using Algorithms 4.1 and 4.2 for the composite hypothesis. In order to examine the possible effects of different functional coefficients  $\boldsymbol{\rho}$  and underlying processes for  $\mathbf{X}$ , we considered nine possible scenarios combining both factors. The detailed description of these scenarios is given in the supplement, while a coarse-grained graphical idea can be obtained from Figure 1.

The different data generating processes are encoded as follows. For the  $k$ -th simulation scenario  $Sk$ , with functional coefficient  $\boldsymbol{\rho}_k$ , the deviation from  $H_0$  is measured by a parameter  $\delta_d$ , with  $\delta_0 = 0$  and  $\delta_d > 0$  for  $d = 1, 2$ . Then, with  $H_{k,d}$  we denote the data generation from

$$Y = \langle \mathbf{X}, \boldsymbol{\rho}_k \rangle + \delta_d \Delta_{\lceil \frac{k}{4} \rceil}(\mathbf{X}) + \varepsilon,$$

where the deviations from the linear model are constructed by including the non-linear terms  $\Delta_1(\mathbf{X}) := \|\mathbf{X}\|$ ,  $\Delta_2(\mathbf{X}) := 25 \int_0^1 \int_0^1 \sin(2\pi ts) s(1-s)t(1-t) \mathbf{X}(s) \mathbf{X}(t) ds dt$  and  $\Delta_3(\mathbf{X}) := \langle e^{-\mathbf{X}}, \mathbf{X}^2 \rangle$ . The error  $\varepsilon$  is distributed as a  $\mathcal{N}(0, \sigma^2)$ , where  $\sigma^2$  was chosen such that, under  $H_0$ ,  $R^2 = \frac{\text{Var}[\langle \mathbf{X}, \boldsymbol{\rho} \rangle]}{\text{Var}[\langle \mathbf{X}, \boldsymbol{\rho} \rangle] + \sigma^2} = 0.95$ .  $d_n$  is chosen automatically by SICc.

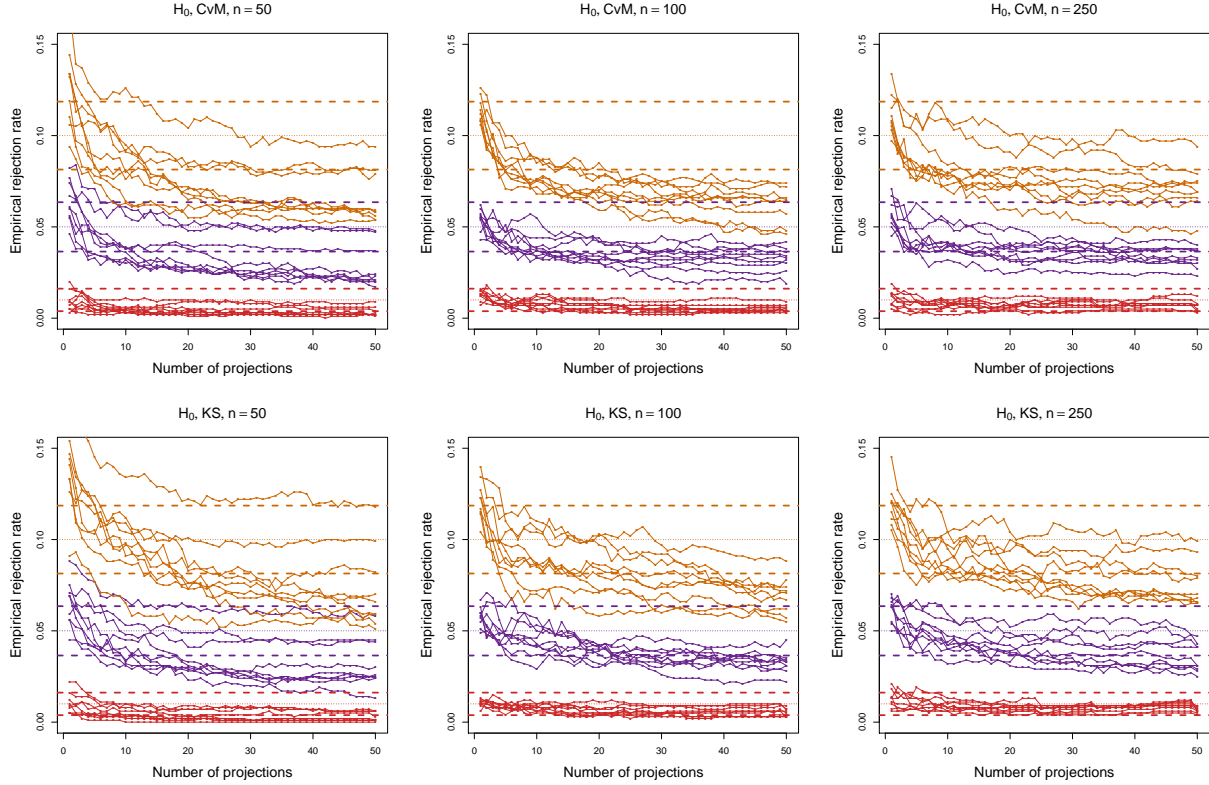


Figure 2: Empirical sizes of the CvM (upper row) and KS (lower row) tests for  $Sk$ ,  $k = 1, \dots, 9$ , depending on the number of projections  $K = 1, \dots, 50$  and for sample sizes  $n = 50, 100, 250$  (from left to right). The empirical sizes associated to the significance levels  $\alpha = 0.01, 0.05, 0.10$  are coded in red, purple and orange, respectively. Dashed thick lines represent the asymptotic 95% confidence interval for the proportion  $\alpha$  obtained from  $M$  replicates.

We explore first the dependence of the tests with respect to the number of projections  $K$ , which are obtained from the data-driven Gaussian processes described in Section 4 (see the supplement for other data generating processes). Figure 2 shows the empirical level for each scenario, based on  $M = 1000$  Monte Carlo trials and  $B = 10000$  bootstrap replicates. There is a clear L-shaped

pattern of the empirical rejection rate curves, which is produced by the conservativeness of the FDR correction – which under  $H_0$  ensures that the rejection rate is *at most*  $\alpha$  – on dealing with the highly-dependent projected tests. For small  $K$ 's, both tests calibrate correctly the three levels for different sample sizes, with the exception of  $n = 50$  and  $\alpha = 0.10$ , for which the tests have a significant over-rejection of the null. For moderate to large  $K$ 's, the empirical rejection rates decrease and stabilize below  $\alpha$ , resulting in a violation of the confidence intervals in a significant number of times, specially for  $\alpha = 0.10$ . Figure 3 shows that the empirical powers with respect to  $K$  are almost constant or have mild decrements, except for certain bumps at lower values of  $K$  that provide a power gain. Both facts point towards choosing the number of projections  $K$  to be relatively small,  $K \in [1, 10]$ , in order to make a reasonable compromise between correct calibration and power. In addition to the computational expediency that a small  $K$  yields, it also avoids requiring a large  $B$  to estimate properly the FDR  $p$ -values, provided that the FDR correction requires an finer precision in the discretization of the  $p$ -values for larger  $K$  (see supplement).

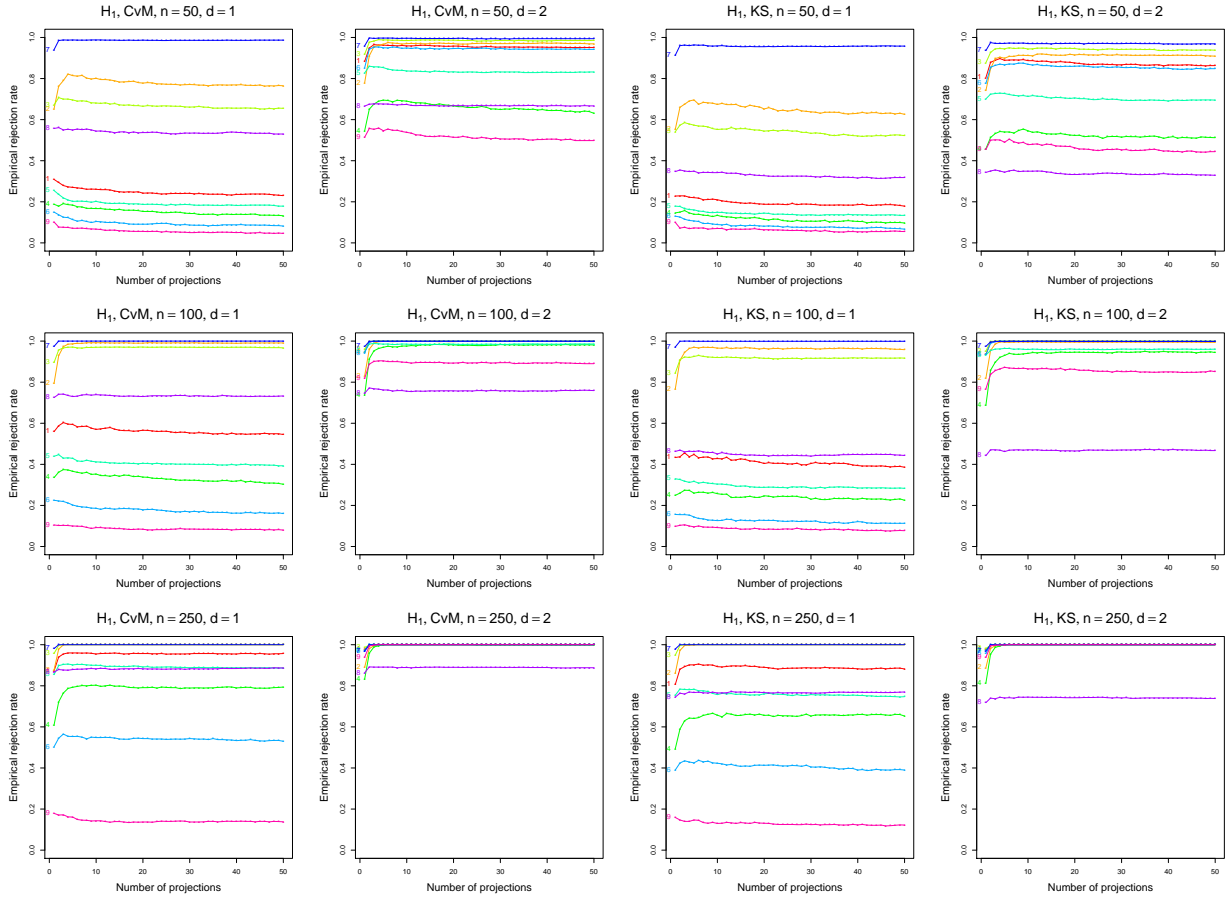


Figure 3: Empirical powers of the CvM (first two columns) and KS (last two) tests for  $Sk$ ,  $k = 1, \dots, 9$ , depending on the number of projections  $K = 1, \dots, 50$ . Odd columns correspond to the deviation index  $d = 1$ , while even account for  $d = 2$ . The significance level is  $\alpha = 0.05$  and the sample sizes are  $n = 50, 100, 250$  (rows, from up to down).

The tests based on the KS and CvM norms are compared with the test presented in García-Portugués et al. (2014) (denoted by PCvM), available in the R package `fda.usc` (Febrero-Bande and Oviedo de la Fuente, 2017), and whose test statistic can be regarded as the average of projected CvM statistics. The test was run with the same FPC estimation used in the new tests, the same number of components  $d_n$  and  $B = 1000$ . Table 1 presents the empirical rejection rates of the different simulation scenarios with  $K = 1, 5, 10$  for KS and CvM tests. The results show two consistent

patterns. First, the CvM test consistently dominates over the KS test, with only two exceptions:  $H_{7,2}$  with  $n = 100$ , and  $H_{7,1}$  with  $n = 50$  (see supplement for the latter). Second, PCvM tends to have a larger power than CvM for most of the situations, specially for small sample sizes and mild deviations. As an illustration, for  $n = 50$  the average relative loss in the empirical power for CvM<sub>5</sub> with respect to PCvM is 14.8% ( $d = 1$ ) and 4.8% ( $d = 2$ ). For  $n = 100$ , the losses drop to 10.8% and 0.9%, and for  $n = 250$ , to 5.2% and 0.1%. The drop in performance for CvM with respect to PCvM is expected due to the construction of CvM, which opts for exploring a set of random directions instead of averaging uniformly distributed finite-dimensional directions, as PCvM does. This also yields one the strongest points of the CvM test, which is its relatively short running times, specially for large  $n$ . Not surprisingly, the number of evaluations performed for computing the statistic is  $\mathcal{O}(n)$ , a notable reduction from PCvM's  $\mathcal{O}((n^3 - n^2)/2)$ . Also, the memory requirement for CvM is  $\mathcal{O}(n)$ , instead of PCvM's  $\mathcal{O}((n^2 - n - 2)/2)$ . The running times in Figure 4 evidence this improvement.

$H_{k,\delta}$	$n = 100$							$n = 250$						
	CvM <sub>1</sub>	CvM <sub>5</sub>	CvM <sub>10</sub>	KS <sub>1</sub>	KS <sub>5</sub>	KS <sub>10</sub>	PCvM	CvM <sub>1</sub>	CvM <sub>5</sub>	CvM <sub>10</sub>	KS <sub>1</sub>	KS <sub>5</sub>	KS <sub>10</sub>	PCvM
$H_{1,0}$	0.062	0.039	0.033	0.051	0.037	0.029	0.046	0.057	0.038	0.032	0.055	0.045	0.044	0.051
$H_{2,0}$	0.060	0.047	0.044	0.058	0.063	0.050	0.068	0.056	0.035	0.040	0.070	0.047	0.040	0.062
$H_{3,0}$	0.043	0.035	0.033	0.049	0.043	0.043	0.046	0.071	0.047	0.041	0.063	0.051	0.047	0.055
$H_{4,0}$	0.054	0.045	0.038	0.063	0.050	0.053	0.032	0.067	0.063	0.052	0.066	0.067	0.059	0.034
$H_{5,0}$	0.057	0.042	0.037	0.060	0.048	0.041	0.050	0.052	0.040	0.041	0.054	0.046	0.040	0.052
$H_{6,0}$	0.047	0.047	0.055	0.067	0.047	0.057	0.040	0.045	0.037	0.042	0.063	0.045	0.042	0.039
$H_{7,0}$	0.056	0.038	0.034	0.055	0.048	0.042	0.056	0.050	0.037	0.037	0.050	0.038	0.032	0.046
$H_{8,0}$	0.055	0.041	0.037	0.060	0.040	0.037	0.041	0.055	0.043	0.039	0.064	0.050	0.042	0.053
$H_{9,0}$	0.055	0.051	0.039	0.057	0.048	0.046	0.063	0.062	0.049	0.053	0.068	0.062	0.056	0.069
$H_{1,1}$	0.559	0.594	0.570	0.433	0.448	0.428	0.717	0.872	0.960	0.955	0.806	0.902	0.893	0.986
$H_{2,1}$	0.895	0.971	0.970	0.844	0.925	0.920	0.991	0.957	1.000	1.000	0.949	1.000	1.000	1.000
$H_{3,1}$	0.225	0.202	0.186	0.156	0.144	0.126	0.261	0.503	0.553	0.548	0.390	0.425	0.423	0.658
$H_{4,1}$	0.335	0.366	0.353	0.249	0.260	0.257	0.417	0.608	0.793	0.803	0.493	0.641	0.657	0.874
$H_{5,1}$	0.439	0.429	0.412	0.330	0.317	0.304	0.487	0.854	0.902	0.900	0.754	0.783	0.757	0.932
$H_{6,1}$	0.793	0.987	0.992	0.766	0.970	0.966	0.995	0.876	1.000	1.000	0.860	0.999	1.000	1.000
$H_{7,1}$	0.974	1.000	1.000	0.971	0.999	0.999	1.000	0.981	1.000	1.000	0.979	1.000	1.000	1.000
$H_{8,1}$	0.725	0.730	0.739	0.463	0.462	0.451	0.728	0.867	0.879	0.883	0.745	0.767	0.765	0.867
$H_{9,1}$	0.105	0.100	0.092	0.098	0.096	0.092	0.127	0.179	0.161	0.143	0.159	0.147	0.130	0.200
$H_{1,2}$	0.947	1.000	1.000	0.934	0.999	1.000	1.000	0.978	1.000	1.000	0.976	1.000	1.000	1.000
$H_{2,2}$	0.960	1.000	1.000	0.951	1.000	1.000	1.000	0.987	1.000	1.000	0.982	1.000	1.000	1.000
$H_{3,2}$	0.943	1.000	1.000	0.934	1.000	0.999	1.000	0.968	1.000	1.000	0.957	1.000	1.000	1.000
$H_{4,2}$	0.736	0.970	0.976	0.687	0.930	0.940	0.989	0.832	0.999	1.000	0.813	0.999	1.000	0.999
$H_{5,2}$	0.948	0.985	0.985	0.932	0.964	0.959	0.987	0.973	1.000	1.000	0.973	1.000	1.000	1.000
$H_{6,2}$	0.830	0.998	0.998	0.819	0.995	0.995	0.999	0.890	0.999	1.000	0.888	0.999	1.000	1.000
$H_{7,2}$	0.973	0.999	0.999	0.974	0.997	0.999	0.999	0.970	1.000	1.000	0.969	1.000	1.000	1.000
$H_{8,2}$	0.746	0.762	0.756	0.442	0.462	0.471	0.779	0.863	0.891	0.889	0.718	0.741	0.745	0.893
$H_{9,2}$	0.818	0.904	0.896	0.767	0.872	0.865	0.928	0.939	1.000	1.000	0.938	0.999	1.000	1.000

Table 1: Empirical sizes and powers of the CvM, KS and PCvM tests with  $\alpha = 0.05$ , sample sizes  $n = 100, 250$  and estimation of  $\boldsymbol{\rho}$  by data-driven FPC ( $d_n$  chosen by SICc). KS and CvM tests are shown with 1, 5 and 10 projections.

The new tests were also applied to the two data applications described in García-Portugués et al. (2014), yielding similar conclusions. First, for the Tecator data set (prediction of fat content using absorbances curves) the  $p$ -values obtained for 5 and 10 projections were 0.01 and 0.02, respectively, for both CvM and KS (PCvM:  $p$ -value = 0.004). Employing the first derivatives of the absorbance curves gave  $p$ -values = 0.010, 0.020 (CvM), and 0.025, 0.010 (KS). With the second derivatives, all

the  $p$ -values were null. In addition, the tests for  $H_0 : \boldsymbol{\rho} = \mathbf{0}$  had null  $p$ -values for all projections and kinds of tests. As a consequence, we conclude that there are strong evidences against the FLM and that there is a significant relation between the fat content and the absorbances curves. Second, for the AEMET example (yearly wind speed explained by temperature profiles in 73 Spanish weather stations) the test for 1 – 10 projections provides  $p$ -values in the range  $[0.28, 0.69]$  for CvM and in the range  $[0.26, 0.49]$  for KS (PCvM:  $p$ -value = 0.121), concluding a non rejection of the linear model. The tests for  $H_0 : \boldsymbol{\rho} = \mathbf{0}$  gave null  $p$ -values. Hence, we conclude that there is no evidence against the FLM and that the effect of the covariate on the response is significant.

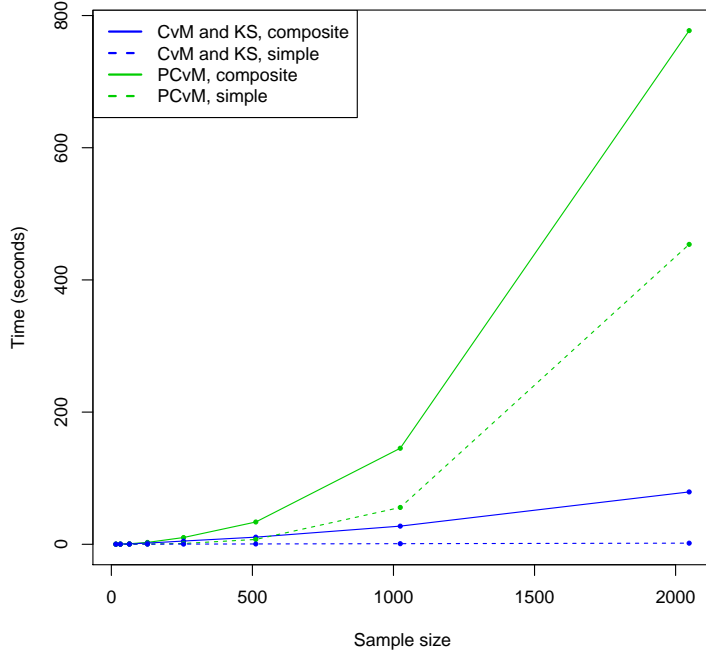


Figure 4: Running times for the tests CvM and KS (the computation of both tests is done in the same routine) and PCvM, for the composite and simple hypotheses. The tests were averaged over  $M = 100$  trials and calibrated with  $B = 1000$ . The sample sizes are  $n = 2^k$ ,  $k = 2, \dots, 11$  and the number of projections considered is  $K = 5$ . Times were measured on a 2.53 GHz core. All the tests were implemented in R, interfacing FORTRAN for the computation of the statistics.

## 6 Discussion

We have presented a new way of building goodness-of-fit tests for regression models with functional covariate employing random projections. The methodology was illustrated using randomly projected empirical processes, which provided root- $n$  consistent tests for testing functional linearity. The calibration of the test was done by a wild bootstrap resampling and the FDR was used to combine  $K$   $p$ -values coming from different projections to account for a higher power. The empirical analysis of the tests, conducted in a fully data-driven way, showed that CvM yields higher powers than KS and that a selection of  $K \in [1, 10]$ , in particular  $K = 5$ , is a reasonable compromise between respecting size and increasing power. There is still a price to pay in terms of a moderate loss of power with respect to the PCvM test, which averages across a set of uniformly distributed finite-dimensional directions. However, the reduction in computational complexity of the new proposals is more than notable.

We conclude the paper by sketching some promising extensions of the methodology for the testing of complexer models involving functional covariates:

- a) Testing the significance of the functional covariate of  $(\mathbf{X}, \mathbf{W}) \in \mathcal{H} \times \mathbb{R}^q$  in the functional partially linear model (Aneiros-Pérez and Vieu, 2006)  $Y = m(\mathbf{X}) + \mathbf{W}^T \boldsymbol{\beta} + \varepsilon$ . The process to be considered from a sample  $\{(\mathbf{X}_i, \mathbf{W}_i, Y_i)\}_{i=1}^n$  and an estimator  $\hat{\boldsymbol{\beta}}$  such that  $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$  is  $n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^h \leq x\}} (Y_i - \mathbf{W}_i^T \hat{\boldsymbol{\beta}})$ .
- b) Testing a functional quadratic regression model (Horvath and Reeder, 2013).
- c) Testing the significance of a functional linear model with functional response:  $H_0 : \mathbb{E}[\mathbf{Y}|\mathbf{X}] = \mathbf{0}$ , where now  $(\mathbf{X}, \mathbf{Y}) \in \mathcal{H}_1 \times \mathcal{H}_2$  and the associated empirical process is  $n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{h_1} \leq x\}} \mathbf{Y}_i^{h_2}$ .

## Software availability

The R package `rp.flm.test`, openly available at <https://github.com/egarpor/rp.flm.test>, contains the implementation of both tests and allows to reproduce the simulation study and the data applications. The main function, `rp.flm.test`, is also included in the R package `fda.usc` since version 1.3.1.

## Acknowledgements

The authors acknowledge the support of projects MTM2013–41383–P, MTM2008–06067, and MTM–2014–56235–C2–2–P, from the Spanish Ministry of Science and Innovation and FEDER, project 10MDS207015PR from Dirección Xeral de I+D, Xunta de Galicia, and IAP network StUDyS, from Belgian Science Policy. Work of E. García-Portugués has been supported by FPU grant AP2010–0957 from the Spanish Ministry of Education, and the Dynamical Systems Interdisciplinary Network, University of Copenhagen.

## Supplement

Two extra appendices are included as supplementary material, containing the proofs of the technical lemmas and further results for the simulation study.

## A Proofs of the main results

### A.1 Hypothesis projection

*Proof of Proposition 2.1.* We denote by  $\mathbf{X}_p$  to both the vectors  $(X_1, \dots, X_p)$  and  $(X_1, \dots, X_p, 0, \dots)$  containing the first  $p$  coefficients of  $\mathbf{X}$  in an orthonormal basis of  $\mathcal{H}$ . We prove first the result for the finite subspace of  $\mathcal{H}$  spanned by the first  $p$  elements of the orthonormal basis. We have to show that

$$\mathbb{E}[Y|\mathbf{X}_p] = 0 \text{ a.s.} \iff \mathbb{E}[Y|\langle \mathbf{X}_p, \mathbf{h} \rangle] = 0 \text{ a.s., for every } \mathbf{h} \in \mathcal{H}. \quad (7)$$

To prove this, we make use of Theorem 1 in Bierens (1982) which states that if  $\mathbf{V}$  and  $\mathbf{Z}$  are two  $\mathbb{R}^p$ -valued random vectors, then

$$\mathbb{E}[\mathbf{V}|\mathbf{Z}] = \mathbf{0} \text{ a.s.} \iff \mathbb{E}[\mathbf{V}e^{i\langle \mathbf{t}, \mathbf{Z} \rangle}] = \mathbf{0}, \text{ for every } \mathbf{t} \in \mathbb{R}^p. \quad (8)$$

Assume that  $\mathbb{E}[Y|\mathbf{X}_p] = 0$  and let  $\mathbf{h} \in \mathcal{H}$ . Since  $\sigma(\langle \mathbf{X}_p, \mathbf{h} \rangle) \subset \sigma(\mathbf{X}_p)$ , we have that  $\mathbb{E}[Y|\langle \mathbf{X}_p, \mathbf{h} \rangle] = \mathbb{E}[\mathbb{E}[Y|\mathbf{X}_p]|\langle \mathbf{X}_p, \mathbf{h} \rangle] = 0$  a.s., which shows the *if* part. To obtain the *only if* part, let  $\mathbf{h} \in \mathcal{H}$ , and

compute  $\mathbb{E}[Y e^{it\langle \mathbf{X}_p, \mathbf{h} \rangle}] = \mathbb{E}[\mathbb{E}[Y | \langle \mathbf{X}_p, \mathbf{h} \rangle] e^{it\langle \mathbf{X}_p, \mathbf{h} \rangle}] = 0$ , for every  $t \in \mathbb{R}$ . Then (7) follows from (8).

Now we are in position to prove the result for  $\mathcal{H}$ . As before, the *if* implication follows from the fact that  $\sigma(\langle \mathbf{X}, \mathbf{h} \rangle) \subset \sigma(\mathbf{X})$ . To prove the inverse, given  $p \in \mathbb{N}$  and  $\mathbf{h} \in \mathcal{H}$ , since  $\mathbf{h}_p \in \mathcal{H}$  and  $\langle \mathbf{X}, \mathbf{h}_p \rangle = \langle \mathbf{X}_p, \mathbf{h} \rangle$ , then  $\sigma(\langle \mathbf{X}_p, \mathbf{h} \rangle) \subset \sigma(\langle \mathbf{X}, \mathbf{h} \rangle)$ , and we have that the assumption implies that  $\mathbb{E}[Y | \langle \mathbf{X}_p, \mathbf{h} \rangle] = 0$  a.s. Thus, from (7), we have that  $\mathbb{E}[Y | \mathbf{X}_p] = 0$  a.s. for every  $p$ , and the result follows from the fact that  $\sigma(\mathbf{X}_p) \uparrow \sigma(\mathbf{X})$  because of the integrability assumption on  $Y$ .  $\square$

*Proof of Lemma 2.3.* From the properties of the conditional expectation, the Cauchy-Schwartz and Jensen inequalities, we have that

$$l_k = \mathbb{E} \left[ \|\mathbf{X}\|^k \mathbb{E}[|Y| | \mathbf{X}] \right] \leq (m_{2k})^{1/2} (\mathbb{E}[Y^2])^{1/2}.$$

thus  $l_k$  is finite. By the convexity of the function  $t \rightarrow t^{(2k+1)/2k}$  and Jensen's inequality,  $m_{2k}^{1/2k} \leq m_{2k+1}^{1/(2k+1)}$ , hence  $\sum_{k=1}^{\infty} m_{2k}^{-1/2k} = \infty$ .  $\square$

*Proof of Theorem 2.4.* The *only if* part is trivial because  $\sigma(\mathbf{X}^{\mathbf{h}}) \subset \sigma(\mathbf{X})$ , and then  $\mathbb{E}[Y | \mathbf{X}] = 0$  a.s. implies that  $\mu(\mathcal{H}_0) = 1$ . Concerning the *if* part, let us assume that  $\mu(\mathcal{H}_0) > 0$ . From the assumptions we have that  $\mathbb{E}[|Y| | \mathbf{X}] < \infty$  and, if we take  $\mathbf{h} \in \mathcal{H}_0$ , then

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y | \mathbf{X}^{\mathbf{h}}]] = 0. \quad (9)$$

Let us assume that  $\mathbb{E}[Y | \mathbf{X}]$  is not zero a.s. Then, the random variables

$$\begin{aligned} \Phi^+(\mathbf{X}) &:= (\mathbb{E}[Y | \mathbf{X}])^+ = \max\{\mathbb{E}[Y | \mathbf{X}], 0\}, \\ \Phi^-(\mathbf{X}) &:= (\mathbb{E}[Y | \mathbf{X}])^- = \max\{-\mathbb{E}[Y | \mathbf{X}], 0\}, \end{aligned}$$

are integrable and positive with positive probability. Thus, (9) implies that

$$V := \int \Phi^+(\mathbf{x}) dP_{\mathbf{X}}(\mathbf{x}) = \int \Phi^-(\mathbf{x}) dP_{\mathbf{X}}(\mathbf{x}) > 0.$$

Consider now the probability measures  $\nu_{\Phi}^+$  and  $\nu_{\Phi}^-$ , defined on  $\mathcal{H}$  and whose density functions with respect to  $P_{\mathbf{X}}$  are respectively

$$\frac{d\nu_{\Phi}^+}{dP_{\mathbf{X}}}(\mathbf{x}) := V^{-1} \Phi^+(\mathbf{x}) \quad \text{and} \quad \frac{d\nu_{\Phi}^-}{dP_{\mathbf{X}}}(\mathbf{x}) := V^{-1} \Phi^-(\mathbf{x}).$$

For  $k \in \mathbb{N}$ , the moments of  $\nu_{\Phi^+}$  verify (analogously for  $\Phi^-$ )

$$\int \|\mathbf{x}\|^k d\nu_{\Phi^+}(\mathbf{x}) \leq V^{-1} \int \|\mathbf{x}\|^k \mathbb{E}[|Y| | \mathbf{X} = \mathbf{x}] dP_{\mathbf{X}}(\mathbf{x}) = l_k,$$

and then, because of Lemma 2.3, they satisfy (a) in Lemma 2.2. Given  $\mathbf{h} \in \mathcal{H}_0$ , the r.v.  $\mathbf{X}^{\mathbf{h}}$  is  $\mathbf{X}$ -measurable. Thus, a.s.,

$$0 = \mathbb{E}[Y | \mathbf{X}^{\mathbf{h}}] = \mathbb{E}[\mathbb{E}[Y | \mathbf{X}] | \mathbf{X}^{\mathbf{h}}] = \mathbb{E}[\mathbb{E}[Y | \mathbf{X}]^+ | \mathbf{X}^{\mathbf{h}}] - \mathbb{E}[\mathbb{E}[Y | \mathbf{X}]^- | \mathbf{X}^{\mathbf{h}}].$$

From here, it is easy to prove that the marginal distributions of  $\nu_{\Phi}^+$  and  $\nu_{\Phi}^-$  on the one-dimensional subspace generated by  $\mathbf{X}^{\mathbf{h}}$  coincide if  $\mathbf{h} \in \mathcal{H}_0$ . Since  $\mathcal{H}_0$  has a positive  $\mu$ -measure, from Lemma 2.2, we obtain that these probability measures indeed coincide and, as a consequence,  $V^{-1} (\mathbb{E}[Y | \mathbf{X}])^+ = V^{-1} (\mathbb{E}[Y | \mathbf{X}])^-$  a.s., that trivially implies  $\mathbb{E}[Y | \mathbf{X}] = 0$  a.s.  $\square$



## A.2 Testing the linear model

*Proof of Theorem 3.1.* We analyse the asymptotic distribution of the three terms separately by invoking some auxiliary lemmas, whose proofs are in the supplementary material.

The asymptotic distribution of  $T_{n,\mathbf{h}}^1(x)$  follows from Corollary 2.6:  $n^{-1/2}T_{n,\mathbf{h}}^1(x) \xrightarrow{\mathcal{L}} \mathcal{N}(0, K_1(x, x))$ . So, if  $a_n = o(n^{-1/2})$ , then  $a_n T_{n,\mathbf{h}}^1(x) = o_{\mathbb{P}}(1)$ . The following two lemmas give insights into the asymptotic behaviour of  $k_n$  and are required for the analysis of  $T_{n,\mathbf{h}}^2(x)$  and  $T_{n,\mathbf{h}}^3(x)$ .

**Lemma A.1.** *Under A6 and A9,  $k_n^3(\log k_n)^2 = o(n^{1/2})$ .*

**Lemma A.2.** *Under A6 and A9, we have that  $\nu[d_n = k_n] \rightarrow 1$ .*

We employ the decomposition (11) from page 338 in CMS to have

$$\hat{\rho} - \rho = \mathbf{L}_n + \mathbf{Y}_n + \mathbf{S}_n + \mathbf{R}_n, \quad (10)$$

where  $\mathbf{L}_n = \sum_{j=k_n+1}^{\infty} \langle \rho, \mathbf{e}_j \rangle \mathbf{e}_j$ ,  $\mathbf{Y}_n = \sum_{j=1}^{k_n} (\langle \rho, \hat{\mathbf{e}}_j \rangle \hat{\mathbf{e}}_j - \langle \rho, \mathbf{e}_j \rangle \mathbf{e}_j)$ ,  $\mathbf{S}_n = (\Gamma_n^\dagger - \Gamma^\dagger) \mathbf{U}_n$ ,  $\mathbf{R}_n = \Gamma^\dagger \mathbf{U}_n$  and  $\mathbf{U}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \otimes \varepsilon_i$ . The decomposition (11) in CMS contains an extra term,  $\mathbf{T}_n$ , which we can consider null here because Lemma A.2 gives that, asymptotically and with probability one,  $k_n$  can be replaced by  $d_n$  in (5).

We will profusely employ the notation

$$\bar{\mathbf{X}}_{x,\mathbf{h}} := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \mathbf{X}_i.$$

By (10), the term  $T_{n,\mathbf{h}}^2(x)$  can be expressed as

$$T_{n,\mathbf{h}}^2(x) = n \langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{L}_n + \mathbf{S}_n + \mathbf{Y}_n + \mathbf{R}_n \rangle.$$

As a consequence of the following lemmas, we have that  $T_{n,\mathbf{h}}^2(x) = o_{\mathbb{P}}(n^{1/2})$ .

**Lemma A.3.** *Under A3 and A4,  $n^{1/2} \langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{L}_n \rangle = o_{\mathbb{P}}(1)$ .*

**Lemma A.4.** *Under A6, A8 and A9,  $n^{1/2} \langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{R}_n \rangle = o_{\mathbb{P}}(1)$ .*

**Lemma A.5.** *Under A5, A6, A8 and A9,  $n^{1/2} \langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{S}_n \rangle = o_{\mathbb{P}}(1)$ .*

**Lemma A.6.** *Under A4, A6, A8 and A9,  $n^{1/2} \langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{Y}_n \rangle = o_{\mathbb{P}}(1)$ .*

The behaviour of the third term, yielding statement (a), is given by the next lemma.

**Lemma A.7.** *Under A3, A4, A6, A7 and A9,  $n^{-1/2} t_{n,\mathbf{E}_{x,\mathbf{h}}}^{-1} T_{n,\mathbf{h}}^3(x) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{\varepsilon}^2)$ .*

From the above results,  $a_n T_{n,\mathbf{h}}^2(x) = o_{\mathbb{P}}(1)$  for cases (b) and (c),  $T_{n,\mathbf{h}}^3$  is the dominant term in (b) and both  $T_{n,\mathbf{h}}^1$  and  $T_{n,\mathbf{h}}^3$  are dominant in (c).  $\square$

*Proof of Proposition 3.2.* By the definition of  $t_{n,\mathbf{E}_{x,\mathbf{h}}}$  and (4),

$$t_{n,\mathbf{E}_{x,\mathbf{h}}}^2 = \sum_{j=1}^{k_n} \frac{\mathbb{E} \left[ \mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \leq x\}} \langle \mathbf{X}, \mathbf{e}_j \rangle \right]^2}{\lambda_j} = \sum_{j=1}^{k_n} \mathbb{E} \left[ \mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \leq x\}} \xi_j \right]^2 \leq \sum_{j=1}^{k_n} \mathbb{E} [\xi_j^2] = k_n.$$

We assume now that  $\mathbf{X}$  is Gaussian. Obviously, the two-dimensional random vector  $(\xi_j, \mathbf{X}^{\mathbf{h}})$ ,  $j \in \mathbb{N}$ , is centred normal. Moreover, the variance of  $\xi_j$  is one and, if  $h_j := \langle \mathbf{h}, \mathbf{e}_j \rangle$ , then  $\sigma_{\mathbf{h}}^2 = \sum_{j=1}^{\infty} h_j^2 \lambda_j < \infty$  (because  $\sum_{j=1}^{\infty} h_j^2 < \infty$  and A3) and  $\text{Cov}[\xi_j, \mathbf{X}^{\mathbf{h}}] = h_j \lambda_j^{1/2}$ . Notice that, if  $\mathbf{h} \neq \mathbf{0}$ ,  $\sigma_{\mathbf{h}} > 0$  since

$\lambda_j > 0$ , for all  $j \in \mathbb{N}$ . Denoting by  $\phi_{\mathbf{h}}(u, v)$  to the joint density function of  $(\xi_j, \mathbf{X}^{\mathbf{h}})$ , by  $\phi_{\mathbf{h},2}(v)$  to its second marginal, we have that

$$\begin{aligned}\mathbb{E} \left[ \mathbb{1}_{\{\mathbf{X}^{\mathbf{h}} \leq x\}} \xi_j \right] &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} u \mathbb{1}_{\{v \leq x\}} \frac{\phi_{\mathbf{h}}(u, v)}{\phi_{\mathbf{h},2}(v)} du \right) \phi_{\mathbf{h},2}(v) dv \\ &= \int_{-\infty}^x \mathbb{E}[\xi_j | \mathbf{X}^{\mathbf{h}} = v] \phi_{\mathbf{h},2}(v) dv \\ &= \int_{-\infty}^x \frac{h_j \sqrt{\lambda_j} v}{\sigma_{\mathbf{h}}^2} \phi_{\mathbf{h},2}(v) dv = -\frac{h_j \sqrt{\lambda_j}}{\sigma_{\mathbf{h}}} \phi(x/\sigma_{\mathbf{h}}),\end{aligned}$$

where we have employed the properties of the normal. This, the initial development and **A3** give us that

$$t_{n, \mathbf{E}_{x, \mathbf{h}}}^2 = \frac{\phi^2(x/\sigma_{\mathbf{h}})}{\sigma_{\mathbf{h}}^2} \sum_{j=1}^{k_n} h_j^2 \lambda_j \rightarrow \phi^2(x/\sigma_{\mathbf{h}}).$$

□

*Proof of Theorem 3.3.* We first prove (a). The joint asymptotic normality of  $(T_{n, \mathbf{h}}(x_1), \dots, T_{n, \mathbf{h}}(x_k))$  for  $(x_1, \dots, x_k) \in \mathbb{R}^k$  follows by the Cramér-Wold device and the same arguments used in Lemma A.7. Also, in the proof of that lemma is shown that  $n^{1/2} \langle \mathbf{E}_{x, \mathbf{h}}, \mathbf{L}_n + \mathbf{Y}_n + \mathbf{S}_n \rangle = o_{\mathbb{P}}(1)$ . Then, due to (10) and (4),

$$\begin{aligned}T_{n, \mathbf{h}}(x) &= n^{-1/2} (T_{n, \mathbf{h}}^1(x) + T_{n, \mathbf{h}}^3(x)) \\ &= n^{-1/2} \sum_{i=1}^n \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \varepsilon_i + n^{-1/2} \sum_{i=1}^n \langle \mathbf{E}_{x, \mathbf{h}}, \Gamma^{\dagger} \mathbf{X}_i \rangle \varepsilon_i + o_{\mathbb{P}}(1) \\ &= n^{-1/2} \sum_{i=1}^n \{A_x^i + B_x^i\} + o_{\mathbb{P}}(1)\end{aligned}$$

with  $A_x^i := \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} \varepsilon_i$  and  $B_x^i := \sum_{j=1}^{k_n} \langle \mathbf{E}_{x, \mathbf{h}}, \mathbf{e}_j \rangle \lambda_j^{-1/2} \xi_j^i \varepsilon_i$ . Since the  $\mathbf{X}_i$ 's and  $\varepsilon_i$ 's are i.i.d., and  $\mathbb{E}[\varepsilon | \mathbf{X}] = 0$ ,

$$\text{Cov} \left[ n^{-1/2} \sum_{i=1}^n \{A_s^i + B_s^i\}, n^{-1/2} \sum_{i'=1}^n \{A_t^{i'} + B_t^{i'}\} \right] = \mathbb{E} [A_s^1 A_t^1] + \mathbb{E} [A_s^1 B_t^1] + \mathbb{E} [B_s^1 A_t^1] + \mathbb{E} [B_s^1 B_t^1].$$

Applying the tower property with the conditioning variables  $\mathbf{X}^{\mathbf{h}}$  (first expectation) and  $\mathbf{X}$  (second and third), it follows

$$\begin{aligned}\mathbb{E} [A_s^1 A_t^1] &= K_1(s, t), \\ \mathbb{E} [A_s^1 B_t^1] &= \int_{\{\mathbf{x}^{\mathbf{h}} \leq s\}} \mathbb{V}\text{ar} [Y | \mathbf{X} = \mathbf{x}] \langle \mathbf{E}_{t, \mathbf{h}}, \Gamma^{\dagger} \mathbf{x} \rangle dP_{\mathbf{X}}(\mathbf{x}), \\ \mathbb{E} [B_s^1 B_t^1] &= \int \mathbb{V}\text{ar} [Y | \mathbf{X} = \mathbf{x}] \langle \mathbf{E}_{s, \mathbf{h}}, \Gamma^{\dagger} \mathbf{x} \rangle \langle \mathbf{E}_{t, \mathbf{h}}, \Gamma^{\dagger} \mathbf{x} \rangle dP_{\mathbf{X}}(\mathbf{x}).\end{aligned}$$

Since  $\Gamma^{\dagger} \rightarrow \Gamma^{-1}$  in the operator norm  $\|\cdot\|_{\infty}$ , Cauchy-Schwartz and  $\|(\Gamma^{\dagger} - \Gamma^{-1})\mathbf{x}\| \leq \|\Gamma^{\dagger} - \Gamma^{-1}\|_{\infty} \|\mathbf{x}\|$  give that  $\mathbb{E} [A_s^1 B_t^1] - C_1(s, t)$  and  $\mathbb{E} [B_s^1 B_t^1] - C_2(s, t)$  converge to zero. The result follows then from Slutsky's theorem.

We prove now (b). The tightness of  $n^{-1/2} T_{n, \mathbf{h}}^1$  is obtained with the same arguments as in Theorem 1.1 of Stute (1997). For the tightness of  $n^{-1/2} T_{n, \mathbf{h}}^3$ , define

$$\bar{T}_{n, \mathbf{h}}^3(u) := n \langle \mathbb{E} [\mathbb{1}_{\{U_{\mathbf{h}} \leq u\}} \mathbf{X}], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}} \rangle, \quad U_{\mathbf{h}} = F_{\mathbf{h}}(\mathbf{X}^{\mathbf{h}}),$$

as the time-changed version of  $\bar{T}_{n,\mathbf{h}}^3$  by  $F_{\mathbf{h}}$ , i.e.

$$T_{n,\mathbf{h}}^3(x) = \bar{T}_{n,\mathbf{h}}^3(F_{\mathbf{h}}(x)).$$

Consider  $0 \leq u_1 < u < u_2 \leq 1$  and the differences

$$\begin{aligned} n^{-1/2}(\bar{T}_{n,\mathbf{h}}^3(u) - \bar{T}_{n,\mathbf{h}}^3(u_1)) &= n^{1/2} \langle \mathbb{E}[\mathbb{1}_{\{u_1 < U_{\mathbf{h}} \leq u\}} \mathbf{X}], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}} \rangle, \\ n^{-1/2}(\bar{T}_{n,\mathbf{h}}^3(u_2) - \bar{T}_{n,\mathbf{h}}^3(u)) &= n^{1/2} \langle \mathbb{E}[\mathbb{1}_{\{u < U_{\mathbf{h}} \leq u_2\}} \mathbf{X}], \boldsymbol{\rho} - \hat{\boldsymbol{\rho}} \rangle. \end{aligned}$$

Then, by the Cauchy-Schwartz and Jensen inequalities

$$\begin{aligned} &\mathbb{E} \left[ n^{-2} |\bar{T}_{n,\mathbf{h}}^3(u) - \bar{T}_{n,\mathbf{h}}^3(u_1)|^2 |\bar{T}_{n,\mathbf{h}}^3(u_2) - \bar{T}_{n,\mathbf{h}}^3(u)|^2 \right] \\ &\leq n^2 \mathbb{E} \left[ \|\mathbb{E}[\mathbb{1}_{\{u_1 < U_{\mathbf{h}} \leq u\}} \mathbf{X}]\|^2 \|\mathbb{E}[\mathbb{1}_{\{u < U_{\mathbf{h}} \leq u_2\}} \mathbf{X}]\|^4 \|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4 \right] \\ &= n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4] \int \mathbb{E}[\mathbf{X}(t) \mathbb{1}_{\{u_1 < U_{\mathbf{h}} \leq u\}}]^2 dt \int \mathbb{E}[\mathbf{X}(t) \mathbb{1}_{\{u < U_{\mathbf{h}} \leq u_2\}}]^2 dt \\ &\leq n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4] \int \mathbb{E}[\mathbf{X}^2(t) \mathbb{1}_{\{u_1 < U_{\mathbf{h}} \leq u\}}] dt \int \mathbb{E}[\mathbf{X}^2(t) \mathbb{1}_{\{u < U_{\mathbf{h}} \leq u_2\}}] dt \\ &= n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4] [F(u) - F(u_1)] [F(u_2) - F(u)] \\ &\leq n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4] [F(u_2) - F(u_1)]^2 \\ &= [G(u_2) - G(u_1)]^2, \end{aligned}$$

where  $F(u) := \int \mathbb{E}[\mathbf{X}^2(t) \mathbb{1}_{\{U_{\mathbf{h}} \leq u\}}] dt$  and  $G(u) := \sup_n \{n^2 \mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4]\} F(u)$  is a nondecreasing and continuous function on  $[0, 1]$ . This corresponds to employing  $\gamma = 2$  and  $\alpha = 1$  in Theorem 15.6 of Billingsley (1968), which gives the weak convergence of  $n^{-1/2} \bar{T}_{n,\mathbf{h}}^3$  in  $D([0, 1])$  and, as a consequence of the Continuous Mapping Theorem (CMT),  $n^{-1/2} T_{n,\mathbf{h}}^3 \xrightarrow{\mathcal{L}} \mathcal{G}_2$  in  $D(\mathbb{R})$ .

Finally, we prove that  $n^{-1/2} T_{n,\mathbf{h}}^2 \xrightarrow{p} \mathbf{0}$ . Note first that by Cauchy-Schwartz,

$$\sup_{x \in \mathbb{R}} |n^{-1/2} T_{n,\mathbf{h}}^2(x)| \leq \sup_{x \in \mathbb{R}} \|\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}\| n^{1/2} \|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|.$$

Assumption  $\mathbb{E}[\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}\|^4] = \mathcal{O}(n^{-2})$  implies  $\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\| = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ . Besides, the weak law of large numbers in  $\mathcal{H}$  (e.g. Hoffmann-Jorgensen and Pisier (1976)) and **A3** give  $\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}} \xrightarrow{p} \mathbf{0}$  in  $\mathcal{H}$ . Therefore, the CMT yields  $\sup_{x \in \mathbb{R}} |n^{-1/2} T_{n,\mathbf{h}}^2(x)| \xrightarrow{p} 0$  and as a consequence  $n^{-1/2} T_{n,\mathbf{h}}^2 \xrightarrow{p} \mathbf{0}$  in  $D(\mathbb{R})$ .  $\square$

*Proof of Corollary 3.4.*  $\|T_{n,\mathbf{h}}\|_{\text{KS}} \xrightarrow{\mathcal{L}} \|\mathcal{G}_2\|_{\text{KS}}$  follows from the CMT. For the Cramér-von Mises norm, we use

$$\|T_{n,\mathbf{h}}\|_{\text{CvM}} = \int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2 dF_{\mathbf{h}}(x) + \int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2 d(F_{n,\mathbf{h}} - F_{\mathbf{h}})(x) \quad (11)$$

and  $F_{n,\mathbf{h}} - F_{\mathbf{h}} \xrightarrow{p} \mathbf{0}$ . By Slutsky's theorem  $(T_{n,\mathbf{h}}, F_{n,\mathbf{h}} - F_{\mathbf{h}}) \xrightarrow{\mathcal{L}} (\mathcal{G}_2, \mathbf{0})$ . Then, by the CMT,  $\int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2 d(F_{n,\mathbf{h}} - F_{\mathbf{h}})(x) \xrightarrow{\mathcal{L}} \mathbf{0}$  and  $\int_{\mathbb{R}} T_{n,\mathbf{h}}(x)^2 dF_{\mathbf{h}}(x) \xrightarrow{\mathcal{L}} \int_{\mathbb{R}} \mathcal{G}_2(x)^2 dF_{\mathbf{h}}(x)$ , which ends the proof.  $\square$

## References

Aneiros-Pérez, G. and P. Vieu (2006). Semi-functional partial linear regression. *Statist. Probab. Lett.* 76(11), 1102–1110.

- Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 29(4), 1165–1188.
- Bickel, P. J. and M. Rosenblatt (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* 1(6), 1071–1095.
- Bierens, H. J. (1982). Consistent model specification tests. *J. Econometrics* 20(1), 105–134.
- Billingsley, P. (1968). *Convergence of Probability Measures*. New York: John Wiley & Sons.
- Bücher, A., H. Dette, and G. Wieczorek (2011). Testing model assumptions in functional regression models. *J. Multivariate Anal.* 102(10), 1472–1488.
- Cardot, H., F. Ferraty, A. Mas, and P. Sarda (2003). Testing hypotheses in the functional linear model. *Scand. J. Statist.* 30(1), 241–255.
- Cardot, H., F. Ferraty, and P. Sarda (2003). Spline estimators for the functional linear model. *Statist. Sinica* 13(3), 571–592.
- Cardot, H., A. Mas, and P. Sarda (2007). CLT in functional linear regression models. *Probab. Theory Related Fields* 138(3-4), 325–361.
- Chiou, J.-M. and H.-G. Müller (2007). Diagnostics for functional regression via residual processes. *Comput. Statist. Data Anal.* 51(10), 4849–4863.
- Cuesta-Albertos, J. A., E. del Barrio, R. Fraiman, and C. Matrán (2007). The random projection method in goodness of fit for functional data. *Comput. Statist. Data Anal.* 51(10), 4814–4831.
- Cuesta-Albertos, J. A. and M. Febrero-Bande (2010). A simple multiway ANOVA for functional data. *TEST* 19(3), 537–557.
- Cuesta-Albertos, J. A., R. Fraiman, and T. Ransford (2007). A sharp form of the Cramér-Wold theorem. *J. Theoret. Probab.* 20(2), 201–209.
- Delsol, L., F. Ferraty, and P. Vieu (2011a). Structural test in regression on functional variables. *J. Multivariate Anal.* 102(3), 422–447.
- Delsol, L., F. Ferraty, and P. Vieu (2011b). Structural tests in regression on functional variable. In F. Ferraty (Ed.), *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, pp. 77–83. Berlin: Physica.
- Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* 1(2), 279–290.
- Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* 22(6), 1030–1051.
- Febrero-Bande, M. and M. Oviedo de la Fuente (2017). *fda.usc: Functional Data Analysis and Utilities for Statistical Computing (fda.usc)*. URL <http://cran.r-project.org/web/packages/fda.usc/>. R package version 1.3.1.
- Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis*. Springer Series in Statistics. New York: Springer.
- García-Portugués, E., W. González-Manteiga, and M. Febrero-Bande (2014). A goodness-of-fit test for the functional linear model with scalar response. *J. Comput. Graph. Stat.* 23(3), 761–778.

- González-Manteiga, W. and R. M. Crujeiras (2013). An updated review of goodness-of-fit tests for regression models. *Test* 22(3), 361–411.
- Hall, P. and M. Hosseini-Nasab (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 68(1), 109–126.
- Härdle, W. and E. Mammen (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* 21(4), 1926–1947.
- Hilgert, N., A. Mas, and N. Verzelen (2013). Minimax adaptive tests for the functional linear model. *Ann. Statist.* 41(2), 838–869.
- Hoffmann-Jorgensen, J. and G. Pisier (1976, 08). The law of large numbers and the central limit theorem in Banach spaces. *Ann. Probab.* 4(4), 587–599.
- Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data With Applications*. Springer Series in Statistics. New York: Springer.
- Horvath, L. and R. Reeder (2013). A test of significance in functional quadratic regression. *Bernoulli* 19(5A), 2120–2151.
- Kokoszka, P., I. Maslova, J. Sojka, and L. Zhu (2008). Testing for lack of dependence in the functional linear model. *Canad. J. Statist.* 36(2), 207–222.
- Lavergne, P. and V. Patilea (2008). Breaking the curse of dimensionality in nonparametric testing. *J. Econometrics* 143(1), 103–122.
- McQuarrie, A. D. (1999). A small-sample correction for the Schwarz SIC model selection criterion. *Statist. Probab. Lett.* 44(1), 79 – 86.
- Nadaraya, E. (1964). On estimating regression. *Teor. Veroyatnost. i Primenen* 9(1), 157–159.
- Patilea, V., C. Sánchez-Sellero, and M. Saumard (2012). Projection-based nonparametric testing for functional covariate effect. *arXiv:1205.5578*.
- Patilea, V., C. Sánchez-Sellero, and M. Saumard (2016). Testing the predictor effect on a functional response. *J. Amer. Statist. Assoc.* 111(516), 1684–1695.
- Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (second ed.). Springer Series in Statistics. New York: Springer.
- Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* 25(2), 613–641.
- Stute, W., W. González Manteiga, and M. Presedo Quindimil (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* 93(441), 141–149.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* 26, 359–372.
- Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* 75(2), 263–289.

# Supplement to “Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes”

Juan A. Cuesta-Albertos<sup>1</sup>

Eduardo García-Portugués<sup>2,4</sup>  
Wenceslao González-Manteiga<sup>3</sup>

Manuel Febrero-Bande<sup>3</sup>

## Abstract

This supplement is organized as follows. Section B proves the technical lemmas used in the main results of the paper. Section C gives further details about the simulation study and contains extra results omitted in the paper.

**Keywords:** Empirical process; Functional data; Functional linear model; Functional principal components; Goodness-of-fit; Random projections.

## B Proofs of auxiliary lemmas

Some general setting required for the proof of the auxiliary lemmas is introduced as follows. We will use the notation  $\mathbf{X}_{x,\mathbf{h}}^i := \mathbf{X}_i \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} - \mathbf{E}_{x,\mathbf{h}}$ ,  $i = 1, \dots, n$ , and  $\Gamma_z := zI - \Gamma$ . We also consider the linear operator  $\Gamma_z^{-1/2}$  which is defined in CMS as an operator with the same eigenfunctions as  $\Gamma$  and with eigenvalues equal to  $(z - \lambda_j)^{-1/2}$ , where the square root is taken in the complex space. By  $\xi_j^i$ , we refer to the  $j$ -th random coefficient in the decomposition of  $\mathbf{X}_i$ , in (4). We also write  $\mathbf{X}_{x,\mathbf{h}}^i = \sum_{j=1}^{\infty} D_{x,\mathbf{h}}^{i,j} \mathbf{e}_j$  for the expansion of  $\mathbf{X}_{x,\mathbf{h}}^i$  in the basis of the eigenfunctions of  $\Gamma$ , so  $D_{x,\mathbf{h}}^{i,j} = \xi_j^i \mathbb{1}_{\{\mathbf{X}_i^{\mathbf{h}} \leq x\}} - \mathbb{E}[\xi_j^1 \mathbb{1}_{\{\mathbf{X}_1^{\mathbf{h}} \leq x\}}]$ . We make use of the sets  $\mathcal{B}_j$  (defined in page 339 in CMS), which are the oriented circles of the complex plane with centre  $\lambda_j$  and radius  $\delta_j/2$ , and the functions  $G_n(z)$  defined in the page 351 in the same paper. Finally,  $\tilde{f}_n(z) = z^{-1} \mathbb{1}_{\cup_j \mathcal{B}_j}(z)$  are analytic extensions of  $f_n(x) = x^{-1} \mathbb{1}_{\{x \geq c_n\}}$ . In the proofs a general constant  $C$  (independent from  $x$ ) will appear, which it may change from place to place.

The assumptions stated in Subsection 3.1 could be slightly weakened in the case in which  $\lim_n t_{n,\mathbf{E}_{x,\mathbf{h}}} = \infty$ . The analysis of the proofs show how this can be done depending on the speed of convergence of  $t_{n,\mathbf{E}_{x,\mathbf{h}}}$ .

*Proof of Lemma A.1.* **A6** implies that there exists a finite positive number  $C$  such that  $\lambda_n \leq Cn^{-4} \log n$ . If  $i_n = \lfloor n^{1/7} \rfloor$ , we have that

$$\lambda_{i_n} \leq C \frac{\log n}{i_n^4} = o(n^{-1/2}).$$

Then, by the definition of  $k_n$  and **A9**, we have that  $k_n = \mathcal{O}(n^{1/7})$  and, consequently,

$$k_n^6 (\log k_n)^4 = \mathcal{O}\left(n^{6/7} (\log n)^4\right) = o(n).$$

□

<sup>1</sup>Department of Mathematics, Statistics and Computer Science, University of Cantabria (Spain).

<sup>2</sup>Department of Statistics, Carlos III University of Madrid (Spain).

<sup>3</sup>Department of Statistics and Operations Research, University of Santiago de Compostela (Spain).

<sup>4</sup>Corresponding author. e-mail: edgarcia@est-econ.uc3m.es.

*Proof of Lemma A.2.* Let us define the set

$$\mathcal{A}_n := \left\{ \omega \in \Omega : \sup_{j \leq k_n+1} \frac{|\hat{\lambda}_j - \lambda_j|}{\delta_j} < \frac{1}{2} \right\}.$$

In this set, we have that,

$$\hat{\lambda}_{k_n+1} \leq \lambda_{k_n+1} + \frac{1}{2}\delta_{k_n+1} < c_n$$

and, consequently,  $d_n \leq k_n$ . Moreover, from here, if  $\mathcal{A}_n^j = \mathcal{A}_n \cap \{d_n = j\}$ , then  $\mathcal{A}_n = \bigcup_{j=1}^{k_n} \mathcal{A}_n^j$  and, for every  $j = 1, \dots, k_n$ , on  $\mathcal{A}_n^j$ , it happens that

$$\lambda_j + \frac{\delta_j}{2} > \hat{\lambda}_j = \hat{\lambda}_{d_n} \geq c_n,$$

thus  $k_n \geq d_n$  on  $\mathcal{A}_n$  and  $d_n = k_n$  on this set. However, under **A6** and **A9**, Lemma A.1 holds, and, in particular,  $k_n^2 \log k_n = o(n^{1/2})$ . The proof of Lemma 5 in CMS shows that  $\nu[\mathcal{A}_n] \rightarrow 1$ , what ends our proof.  $\square$

*Proof of Lemma A.3.* By the definition of  $\mathbf{L}_n$ , we have that

$$\langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{L}_n \rangle = \sum_{j=k_n+1}^{\infty} \langle \boldsymbol{\rho}, \mathbf{e}_j \rangle \langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{e}_j \rangle,$$

and, then, taking into account that the  $\mathbf{X}_i$ 's are i.i.d.,

$$\begin{aligned} \mathbb{E} [\langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{L}_n \rangle^2] &= \frac{1}{n} \mathbb{E} \left[ \sum_{j=k_n+1}^{\infty} \langle \boldsymbol{\rho}, \mathbf{e}_j \rangle \left\langle \mathbb{1}_{\{\mathbf{X}_1^{\mathbf{h}} \leq x\}} \mathbf{X}_1 - \mathbf{E}_{x,\mathbf{h}}, \mathbf{e}_j \right\rangle \right]^2 \\ &\leq \frac{1}{n} \mathbb{E} \left[ \sum_{j=k_n+1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_j \rangle| \left\| \mathbb{1}_{\{\mathbf{X}_1^{\mathbf{h}} \leq x\}} \mathbf{X}_1 - \mathbf{E}_{x,\mathbf{h}} \right\| \right]^2 \\ &\leq \frac{1}{n} \mathbb{E} [\|\mathbf{X}_1\|^2] \left( \sum_{j=k_n+1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_j \rangle| \right)^2, \end{aligned}$$

where we have used that

$$\mathbb{E} \left[ \left\| \mathbb{1}_{\{\mathbf{X}_1^{\mathbf{h}} \leq x\}} \mathbf{X}_1 - \mathbf{E}_{x,\mathbf{h}} \right\|^2 \right] \leq \mathbb{E} \left[ \left\| \mathbb{1}_{\{\mathbf{X}_1^{\mathbf{h}} \leq x\}} \mathbf{X}_1 \right\|^2 \right] \leq \mathbb{E} [\|\mathbf{X}_1\|^2].$$

Then the result follows from **A3**, **A4** and the Chebyshev's inequality.  $\square$

*Proof of Lemma A.4.* By the definition of  $\mathbf{R}_n$ , we have that  $\langle \mathbf{R}_n, \mathbf{x} \rangle = n^{-1} \sum_{i=1}^n \langle \Gamma^\dagger \mathbf{X}_i, \mathbf{x} \rangle \varepsilon_i$ . Then

$$\langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{R}_n \rangle = \frac{1}{n} \sum_{i=1}^n \langle \Gamma^\dagger \mathbf{X}_i, \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}} \rangle \varepsilon_i.$$

The  $\varepsilon_i$ 's are i.i.d. and independent from the rest of the involved quantities. Thus

$$\begin{aligned} \mathbb{E} [\langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{R}_n \rangle^2] &= \frac{\sigma_\varepsilon^2}{n} \mathbb{E} [\langle \Gamma^\dagger \mathbf{X}_1, \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}} \rangle^2] \\ &= \frac{\sigma_\varepsilon^2}{n^3} \mathbb{E} \left[ \left( \sum_{i'=1}^n \langle \Gamma^\dagger \mathbf{X}_1, \mathbf{X}_{i'} \mathbb{1}_{\{\mathbf{X}_{i'}^{\mathbf{h}} \leq x\}} - \mathbf{E}_{x,\mathbf{h}} \rangle \right)^2 \right] \end{aligned}$$

$$= \frac{\sigma_\varepsilon^2}{n^3} \mathbb{E} \left[ \left( \sum_{i'=1}^n \sum_{j=1}^{k_n} \xi_j^1 \left( \xi_j^{i'} \mathbb{1}_{\{\mathbf{X}_{i'}^h \leq x\}} - \mathbb{E} \left[ \xi_j^1 \mathbb{1}_{\{\mathbf{X}_1^h \leq x\}} \right] \right) \right)^2 \right]. \quad (12)$$

We have to compute the sum of expectations in (12). This sum contains  $n^2 k_n^2$  terms and we analyse their possible behaviours next.

Let  $i', i'' = 1, \dots, n$ . If  $i' \neq i''$ , then either  $i' \neq 1$  or  $i'' \neq 1$ . Assuming the first one holds, we have that  $(\xi_j^{i'} \mathbb{1}_{\{\mathbf{X}_{i'}^h \leq x\}} - \mathbb{E}[\xi_j^1 \mathbb{1}_{\{\mathbf{X}_1^h \leq x\}}])$  is independent of  $(\xi_j^{i''} \mathbb{1}_{\{\mathbf{X}_{i''}^h \leq x\}} - \mathbb{E}[\xi_j^1 \mathbb{1}_{\{\mathbf{X}_1^h \leq x\}}])$ ,  $\xi_j^1$  and  $\xi_j^{i'}$ . Therefore,

$$\mathbb{E} \left[ \xi_j^1 \left( \xi_j^{i'} \mathbb{1}_{\{\mathbf{X}_{i'}^h \leq x\}} - \mathbb{E} \left[ \xi_j^1 \mathbb{1}_{\{\mathbf{X}_1^h \leq x\}} \right] \right) \xi_j^{i''} \left( \xi_j^{i''} \mathbb{1}_{\{\mathbf{X}_{i''}^h \leq x\}} - \mathbb{E} \left[ \xi_j^1 \mathbb{1}_{\{\mathbf{X}_1^h \leq x\}} \right] \right) \right] = 0,$$

and we have  $k_n^2 n(n-1)$  terms in (12) whose expectations are also zero. With respect to the remaining terms, we can elaborate a bit more on (12) to obtain

$$\begin{aligned} \mathbb{E} [\langle \bar{\mathbf{X}}_{x,h} - \mathbf{E}_{x,h}, \mathbf{R}_n \rangle^2] &\leq \frac{\sigma_\varepsilon^2}{n^3} \mathbb{E} \left[ \left( \sum_{i'=1}^n \sum_{j=1}^{k_n} |\xi_j^1| \left( |\xi_j^{i'}| + \mathbb{E} [|\xi_j^1|] \right) \right)^2 \right] \\ &=: \frac{\sigma_\varepsilon^2}{n^3} \mathbb{E} \left[ \left( \sum_{i'=1}^n \sum_{j=1}^{k_n} T(i', j) \right)^2 \right]. \end{aligned}$$

When expanding the square in the last expression, we obtain the following kind of terms, which can be bounded by  $M$  in **A8** as follows: given  $i', i'' = 1, \dots, n$  and  $j, j' = 1, \dots, k_n$ ,  $\mathbb{E} [T(i', j) T(i'', j')]$  is the sum of four terms each one equal to  $\mathbb{E} [|\xi_j^1| |\xi_j^{i'}| |\xi_{j'}^1| |\xi_{j'}^{i''}|]$ ,  $\mathbb{E} [|\xi_j^1| |\xi_j^{i'}| |\xi_{j'}^1|] \mathbb{E} [|\xi_j^1|]$ , or  $\mathbb{E} [|\xi_j^1| |\xi_{j'}^1|] \mathbb{E} [|\xi_j^1|]$ . If we apply Cauchy-Schwartz's and Jensen's inequalities, by **A8**, we have that, for instance,

$$\mathbb{E} [|\xi_j^1| |\xi_{j'}^1|] \mathbb{E} [|\xi_j^1|] \mathbb{E} [|\xi_{j'}^1|] \leq (\mathbb{E} [|\xi_j^1|^4] \mathbb{E} [|\xi_{j'}^1|^4])^{1/4} \mathbb{E} [|\xi_j^1|^4]^{1/4} \mathbb{E} [|\xi_{j'}^1|^4]^{1/4} \leq M.$$

The remaining terms are handled similarly and we can conclude that the non-null terms in (12) are bounded by  $M$ . Thus, we obtain that

$$\mathbb{E} [\langle \bar{\mathbf{X}}_{x,h} - \mathbf{E}_{x,h}, \mathbf{R}_n \rangle^2] \leq \frac{\sigma_\varepsilon^2}{n^3} M (n^2 k_n^2 - k_n^2 n(n-1)) = \frac{\sigma_\varepsilon^2}{n^2} M k_n^2.$$

Consequently, Chebyshev's inequality and Lemma A.1 give the result.  $\square$

*Proof of Lemma A.5.* The proof follows the steps of proof of Proposition 3 in CMS but replacing the  $\mathbf{X}_{n+1}$  which appears there by the difference  $\bar{\mathbf{X}}_{x,h} - \mathbf{E}_{x,h}$ . We have some technical differences because, here, some involved terms are not independent.

The proof in the beginning of Proposition 3 in CMS allows us to conclude that

$$|\langle \bar{\mathbf{X}}_{x,h} - \mathbf{E}_{x,h}, \mathbf{S}_n \rangle| \leq C \sum_{j=1}^{k_n} H_{j,n}, \quad (13)$$

with

$$H_{j,n} \leq C \int_{\mathcal{B}_j} |\tilde{f}_n(z)| \|G_n(z)\| \|\Gamma_z^{-1/2} \mathbf{U}_n\| \|\Gamma_z^{-1/2} (\bar{\mathbf{X}}_{x,h} - \mathbf{E}_{x,h})\| dz.$$

The application of the Cauchy-Schwartz's inequality twice, plus Lemma 3 in CMS and some arguments developed in Proposition 3 in CMS, give that

$$\mathbb{E} [H_{j,n}] \leq C \int_{\mathcal{B}_j} |\tilde{f}_n(z)| (\mathbb{E} [\|G_n(z)\|^2])^{1/2} \left( \mathbb{E} [\|\Gamma_z^{-1/2} \mathbf{U}_n\|^4] \mathbb{E} [\|\Gamma_z^{-1/2} (\bar{\mathbf{X}}_{x,h} - \mathbf{E}_{x,h})\|^4] \right)^{1/4} dz$$



$$\begin{aligned}
&\leq C \text{diam}(\mathcal{B}_j) \frac{j \log j}{\sqrt{n}} \sup_{z \in \mathcal{B}_j} \left\{ |\tilde{f}_n(z)| \left( \mathbb{E} \left[ \|\Gamma_z^{-1/2} \mathbf{U}_n\|^4 \right] \mathbb{E} \left[ \|\Gamma_z^{-1/2} (\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}})\|^4 \right] \right)^{1/4} \right\} \\
&\leq C \frac{j \log j}{\sqrt{n}} \sup_{z \in \mathcal{B}_j} \left\{ \left( \mathbb{E} \left[ \|\Gamma_z^{-1/2} \mathbf{U}_n\|^4 \right] \mathbb{E} \left[ \|\Gamma_z^{-1/2} (\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}})\|^4 \right] \right)^{1/4} \right\}. \tag{14}
\end{aligned}$$

Let us analyse the two expectations included in (14). First, by definition of  $\mathbf{U}_n$ , we have that

$$\mathbb{E} \left[ \|\Gamma_z^{-1/2} \mathbf{U}_n\|^4 \right] = \frac{1}{n^4} \sum_{r,s,r',s'=1}^n \mathbb{E} \left[ \langle \Gamma_z^{-1/2} \mathbf{X}_r, \Gamma_z^{-1/2} \mathbf{X}_s \rangle \langle \Gamma_z^{-1/2} \mathbf{X}_{r'}, \Gamma_z^{-1/2} \mathbf{X}_{s'} \rangle \right] \mathbb{E} [\varepsilon_r \varepsilon_s \varepsilon_{r'} \varepsilon_{s'}].$$

This sum contains  $n^4$  terms. However, the  $\varepsilon$ 's are centred independent variables, so  $\mathbb{E} [\varepsilon_r \varepsilon_s \varepsilon_{r'} \varepsilon_{s'}]$  equals zero unless the vector  $(r, r', s, s')$  contains two pairs of identical components. This only happens at most in  $n + \frac{1}{2} \binom{4}{2} n(n-1) = 3n^2 - 2n$  terms, and, in those cases,  $\mathbb{E} [\varepsilon_r \varepsilon_s \varepsilon_{r'} \varepsilon_{s'}] = 1$ . Let us compute the value of the other involved expectation in those terms. We have that

$$\langle \Gamma_z^{-1/2} \mathbf{X}_r, \Gamma_z^{-1/2} \mathbf{X}_s \rangle = \sum_{i=1}^{\infty} \frac{\lambda_i}{|z - \lambda_i|} \xi_i^r \xi_i^s$$

and, then,

$$\begin{aligned}
\left| \mathbb{E} \left[ \langle \Gamma_z^{-1/2} \mathbf{X}_r, \Gamma_z^{-1/2} \mathbf{X}_s \rangle \langle \Gamma_z^{-1/2} \mathbf{X}_{r'}, \Gamma_z^{-1/2} \mathbf{X}_{s'} \rangle \right] \right| &\leq \sum_{i=1}^{\infty} \sum_{i'=1}^{\infty} \frac{\lambda_i}{|z - \lambda_i|} \frac{\lambda_{i'}}{|z - \lambda_{i'}|} \mathbb{E} \left[ \left| \xi_i^r \xi_i^s \xi_{i'}^{r'} \xi_{i'}^{s'} \right| \right] \\
&\leq M \sum_{i=1}^{\infty} \sum_{i'=1}^{\infty} \frac{\lambda_i}{|z - \lambda_i|} \frac{\lambda_{i'}}{|z - \lambda_{i'}|},
\end{aligned}$$

where we have applied the Cauchy-Schwartz's inequality twice and **A8**. Those findings, replaced in (15) give that

$$\mathbb{E} \left[ \|\Gamma_z^{-1/2} \mathbf{U}_n\|^4 \right] \leq \frac{3n^2 - 2n}{n^4} M \left( \sum_{i=1}^{\infty} \frac{\lambda_i}{|z - \lambda_i|} \right)^2. \tag{15}$$

We analyse the last factor in (14). We have that

$$\Gamma_z^{-1/2} (\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}) = \frac{1}{n} \sum_{r=1}^n \Gamma_z^{-1/2} \mathbf{X}_{x,\mathbf{h}}^r = \frac{1}{n} \sum_{i=1}^{\infty} \sqrt{\frac{\lambda_i}{z - \lambda_i}} \left( \sum_{r=1}^n D_{x,\mathbf{h}}^{r,i} \right) \mathbf{e}_i.$$

Therefore

$$\|\Gamma_z^{-1/2} (\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}})\|^2 = \frac{1}{n^2} \sum_{i=1}^{\infty} \frac{\lambda_i}{|z - \lambda_i|} \left( \sum_{r=1}^n D_{x,\mathbf{h}}^{r,i} \right)^2 \tag{16}$$

and

$$\mathbb{E} \left[ \|\Gamma_z^{-1/2} (\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}})\|^4 \right] = \frac{1}{n^4} \sum_{i,i'=1}^{\infty} \frac{\lambda_i}{|z - \lambda_i|} \frac{\lambda_{i'}}{|z - \lambda_{i'}|} \sum_{r,r',s,s'=1}^n \mathbb{E} \left[ D_{x,\mathbf{h}}^{r,i} D_{x,\mathbf{h}}^{r',i} D_{x,\mathbf{h}}^{s,j'} D_{x,\mathbf{h}}^{s',j'} \right]. \tag{17}$$

We are in a similar situation to the one in (15) and it happens that all expectations here are zero except, at most,  $3n^2 - 2n$  of them. Moreover, those non-null terms can be bounded. Applying Cauchy-Schwartz inequality twice,

$$\begin{aligned}
\left| \mathbb{E} \left[ D_{x,\mathbf{h}}^{r,j} D_{x,\mathbf{h}}^{r',j} D_{x,\mathbf{h}}^{s,j'} D_{x,\mathbf{h}}^{s',j'} \right] \right| &\leq \mathbb{E} \left[ |D_{x,\mathbf{h}}^{r,j} D_{x,\mathbf{h}}^{r',j} D_{x,\mathbf{h}}^{s,j'} D_{x,\mathbf{h}}^{s',j'}| \right] \\
&\leq \left( \mathbb{E} \left[ (D_{x,\mathbf{h}}^{r,j})^4 \right] \mathbb{E} \left[ (D_{x,\mathbf{h}}^{r',j})^4 \right] \mathbb{E} \left[ (D_{x,\mathbf{h}}^{s,j'})^4 \right] \mathbb{E} \left[ (D_{x,\mathbf{h}}^{s',j'})^4 \right] \right)^{1/4}. \tag{18}
\end{aligned}$$

And, given  $j \in \mathbb{N}$  and  $r = 1, \dots, n$ ,

$$\mathbb{E} \left[ (D_{x,\mathbf{h}}^{r,j})^4 \right] = \mathbb{E} \left[ (\xi_j \mathbb{1}_{\{\mathbf{X}_{\mathbf{h}}^r \leq x\}} - \mathbf{E}_{x,\mathbf{h}}^j)^4 \right] \leq \mathbb{E} \left[ (|\xi_j| + \mathbb{E} [|\xi_j|])^4 \right],$$

what gives an order 4 polynomial whose terms are  $\binom{4}{s} \mathbb{E} [|\xi_j|^s] (\mathbb{E} [|\xi_j|])^{4-s}$  for  $s = 0, \dots, 4$ . If we apply to each of them Jensen's inequality and **A8**, we have that

$$\mathbb{E} [|\xi_j|^s] (\mathbb{E} [|\xi_j|])^{4-s} \leq \mathbb{E} [|\xi_j|^4]^{s/4} (\mathbb{E} [|\xi_j|^4])^{(4-s)/4} = \mathbb{E} [|\xi_j|^4] \leq M.$$

This and (18) give  $|\mathbb{E} [D_{x,\mathbf{h}}^{r,j} D_{x,\mathbf{h}}^{r',j} D_{x,\mathbf{h}}^{s,j'} D_{x,\mathbf{h}}^{s',j'}]| \leq 2^4 M$ , and with (17) yields

$$\mathbb{E} \left[ \left\| \Gamma_z^{-1/2} (\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}) \right\|^4 \right] \leq \frac{3n^2 - 2n}{n^4} 2^4 M \left( \sum_{i=1}^{\infty} \frac{\lambda_i}{|z - \lambda_i|} \right)^2.$$

If we replace this bound and (15) in (14), we obtain that

$$\mathbb{E} [H_{j,n}] \leq C \frac{j \log j}{\sqrt{n}} \frac{(3n^2 - 2n)^2}{n^8} \sup_{z \in \mathcal{B}_j} \left\{ \sum_{i=1}^{\infty} \frac{\lambda_i}{|z - \lambda_i|} \right\}^4.$$

**A5** allows to apply Lemmas 1 and 2 in CMS, which give that  $\sup_{z \in \mathcal{B}_j} \left\{ \sum_{i=1}^{\infty} \frac{\lambda_i}{|z - \lambda_i|} \right\} \leq C j \log j$  and, then, that

$$\mathbb{E} [H_{j,n}] \leq C (j \log j)^5 \frac{(3n^2 - 2n)^2}{n^{17/2}}.$$

This, and (13) give

$$\sqrt{n} \mathbb{E} [|\langle \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}, \mathbf{S}_n \rangle|] \leq C \frac{(3n^2 - 2n)^2}{n^8} \sum_{j=1}^{k_n} (j \log j)^5 \leq C \frac{(3n^2 - 2n)^2}{n^8} k_n^6 (\log k_n)^5 = o(n^{-2})$$

because of **A6**, which proves the lemma.  $\square$

*Proof of Lemma A.6.* This term is handled following the scheme in Proposition 2, page 346, in CMS. Using the notation in that proposition, we have that  $\mathbf{Y}_n = \mathcal{S}_n \boldsymbol{\rho} + \mathcal{R}_n \boldsymbol{\rho}$ . The Cauchy-Schwartz inequality gives

$$\mathbb{E} [|\langle \mathcal{S}_n \boldsymbol{\rho}, \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}} \rangle|] \leq \mathbb{E} [\|\mathcal{S}_n \boldsymbol{\rho}\| \|\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}\|] \leq (\mathbb{E} [\|\mathcal{S}_n \boldsymbol{\rho}\|^2] \mathbb{E} [\|\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}\|^2])^{1/2}. \quad (19)$$

On the other hand, it happens that

$$\begin{aligned} \mathbb{E} [\|\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}}\|^2] &\leq \frac{1}{n^2} \mathbb{E} \left[ \left\| \sum_{r=1}^n \mathbf{X}_{x,\mathbf{h}}^r \right\|^2 \right] = \frac{1}{n^2} \mathbb{E} \left[ \left\| \sum_{l=1}^{\infty} \left( \sum_{r=1}^n D_{x,\mathbf{h}}^{r,l} \right) \mathbf{e}_l \right\|^2 \right] \\ &= \frac{1}{n^2} \sum_{l=1}^{\infty} \mathbb{E} \left[ \left( \sum_{r=1}^n D_{x,\mathbf{h}}^{r,l} \right)^2 \right] = \frac{1}{n^2} \sum_{l=1}^{\infty} \sum_{r=1}^n \mathbb{E} \left[ (D_{x,\mathbf{h}}^{r,l})^2 \right] \\ &= \frac{1}{n} \sum_{l=1}^{\infty} \mathbb{E} \left[ (D_{x,\mathbf{h}}^{1,l})^2 \right] \leq \frac{1}{n} \sum_{l=1}^{\infty} \mathbb{E} [\xi_l^2] = \frac{1}{n} \sum_{l=1}^{\infty} \lambda_l. \end{aligned}$$

From here and (19) we have

$$\mathbb{E} [|\langle \mathcal{S}_n \boldsymbol{\rho}, \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}} \rangle|] \leq \frac{C}{\sqrt{n}} (\mathbb{E} [\|\mathcal{S}_n \boldsymbol{\rho}\|^2])^{1/2} = \frac{C}{\sqrt{n}} \left( \sum_{l=1}^{\infty} \mathbb{E} [\langle \mathcal{S}_n \boldsymbol{\rho}, \mathbf{e}_l \rangle^2] \right)^{1/2}. \quad (20)$$

In pages 347 and 348 in CMS a reasoning is developed giving the next two bounds:

$$\mathbb{E} [\langle \mathcal{S}_n \boldsymbol{\rho}, \mathbf{e}_l \rangle^2] \leq \begin{cases} \frac{M}{n} \left( \sum_{l'=k_n+1}^{\infty} \langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_{l'} - \lambda_l} \right)^2, & \text{if } l \leq k_n, \\ \frac{M}{n} \left( \sum_{l'=1}^{k_n} \langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_{l'} - \lambda_l} \right)^2, & \text{if } l > k_n. \end{cases}$$

This, and (20) give

$$n \left( \mathbb{E} [|\langle \mathcal{S}_n \boldsymbol{\rho}, \bar{\mathbf{X}}_{x, \mathbf{h}} - \mathbf{E}_{x, \mathbf{h}} \rangle|] \right)^2 \leq \frac{C}{n} \sum_{l=1}^{k_n} \left( \sum_{l'=k_n+1}^{\infty} \langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_{l'} - \lambda_l} \right)^2 \quad (21)$$

$$+ \frac{C}{n} \sum_{l=k_n+1}^{\infty} \left( \sum_{l'=1}^{k_n} \langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_{l'} - \lambda_l} \right)^2. \quad (22)$$

Lemma 1 in CMS applied to the term in (21) leads to

$$\begin{aligned} \frac{C}{n} \sum_{l=1}^{k_n} \left( \sum_{l'=k_n+1}^{\infty} \langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_{l'} - \lambda_l} \right)^2 &\leq \frac{C}{n} \sum_{l=1}^{k_n} \left( \sum_{l'=k_n+1}^{\infty} \sqrt{\frac{\lambda_{l'}}{\lambda_l}} \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{1 - \frac{l}{l'}} \right)^2 \\ &\leq \frac{C}{n} \sum_{l=1}^{k_n} \left( \sum_{l'=k_n+1}^{\infty} \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{1 - \frac{l}{l'}} \right)^2 \\ &\leq \frac{2C}{n} \sum_{l=1}^{k_n} \left( \sum_{l'=k_n+1}^{k_n+h_n} \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{1 - \frac{1}{l'}} \right)^2 \\ &\quad + \frac{2C}{n} \sum_{l=1}^{k_n} \left( \sum_{l'=k_n+h_n+1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \frac{1}{1 - \frac{1}{l'}} \right)^2, \end{aligned}$$

where  $h_n = \lfloor \sqrt{\frac{k_n}{\log k_n}} \rfloor$ . From this definition we obtain that the second term satisfies

$$\begin{aligned} \frac{2C}{n} \sum_{l=1}^{k_n} \left( \sum_{l'=k_n+h_n+1}^{\infty} \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{1 - \frac{1}{l'}} \right)^2 &\leq \frac{2C}{n} \sum_{l=1}^{k_n} \left( \sum_{l'=k_n+h_n+1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| (1 + \sqrt{k_n \log k_n}) \right)^2 \\ &\leq \frac{8C}{n} \sum_{l=1}^{k_n} k_n \log k_n \left( \sum_{l'=k_n+1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \right)^2 \\ &\leq \frac{8C}{n} k_n^2 \log k_n \left( \sum_{l'=k_n+1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \right)^2 \rightarrow 0, \end{aligned}$$

where the convergence follows from **A4** and Lemma A.1. On the other hand, the first term verifies

$$\begin{aligned} \frac{2C}{n} \sum_{l=1}^{k_n} \left( \sum_{l'=k_n+1}^{k_n+h_n} \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{1 - \frac{1}{l'}} \right)^2 &\leq \frac{2C}{n} k_n h_n^2 \max_{\substack{k_n < l' \leq k_n+h_n \\ 1 \leq l \leq k_n}} \left\{ \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{\left| 1 - \frac{1}{l'} \right|} \right\}^2 \\ &\leq \frac{2C}{n} k_n \frac{k_n}{\log k_n} (k_n + h_n)^2 \max_{k_n < l'} \{ |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|^2 \} \end{aligned}$$

$$= 8C \frac{k_n^4}{n \log k_n} \max_{k_n < l'} \{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|^2\} \rightarrow 0$$

because of **A4** and Lemma A.1.

Then, the term in (21) converges to zero. Let us analyse the term in (22). Similarly as before, Lemma 1 in CMS gives

$$\begin{aligned} \frac{C}{n} \sum_{l=k_n+1}^{\infty} \left( \sum_{l'=1}^{k_n} \langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_{l'} - \lambda_l} \right)^2 &\leq \frac{C}{n} \sum_{l=k_n+1}^{\infty} \left( \sum_{l'=1}^{k_n} \sqrt{\frac{\lambda_l}{\lambda_{l'}}} \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{1 - \frac{l'}{l}} \right)^2 \\ &\leq \frac{C}{n} \frac{1}{\lambda_{k_n}} \sum_{l=k_n+1}^{\infty} \lambda_l \left( \sum_{l'=1}^{k_n} \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{1 - \frac{l'}{l}} \right)^2 \\ &\leq \frac{2C}{n} \frac{1}{\lambda_{k_n}} \sum_{l=k_n+1}^{k_n+h_n} \lambda_l \left( \sum_{l'=1}^{k_n} \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{1 - \frac{l'}{l}} \right)^2 \\ &\quad + \frac{2C}{n} \frac{1}{\lambda_{k_n}} \sum_{l=k_n+h_n+1}^{\infty} \lambda_l \left( \sum_{l'=1}^{k_n} \frac{|\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle|}{1 - \frac{l'}{l}} \right)^2. \end{aligned} \quad (23)$$

We have that

$$\begin{aligned} \sum_{l=k_n+h_n+1}^{\infty} \lambda_l \left( \sum_{l'=1}^{k_n} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \frac{1}{1 - \frac{l'}{l}} \right)^2 &\leq \sum_{l=k_n+h_n+1}^{\infty} \lambda_l \left( \sum_{l'=1}^{k_n} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \left( 1 + \frac{k_n}{h_n} \right) \right)^2 \\ &\leq 4k_n \log k_n \left( \sum_{l'=1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \right)^2 \sum_{l=k_n+h_n+1}^{\infty} \lambda_l \end{aligned}$$

and that

$$\begin{aligned} \sum_{l=k_n+1}^{k_n+h_n} \lambda_l \left( \sum_{l'=1}^{k_n} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \frac{1}{1 - \frac{l'}{l}} \right)^2 &\leq \sum_{l=k_n+1}^{k_n+h_n} \lambda_l \left( \sum_{l'=1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \right)^2 \max_{\substack{k_n < l < k_n+h_n \\ l' \leq k_n}} \left\{ \frac{1}{1 - \frac{l'}{l}} \right\}^2 \\ &\leq (k_n + h_n)^2 \left( \sum_{l'=1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \right)^2 \sum_{l=k_n+1}^{k_n+h_n} \lambda_l \\ &\leq 4k_n^2 \left( \sum_{l'=1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \right)^2 \sum_{l=k_n+1}^{k_n+h_n} \lambda_l. \end{aligned}$$

Replacing the last two inequalities in (23), we obtain that

$$\begin{aligned} \frac{C}{n} \sum_{l=k_n+1}^{\infty} \left( \sum_{l'=1}^{k_n} \langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle \frac{\sqrt{\lambda_l \lambda_{l'}}}{\lambda_{l'} - \lambda_l} \right)^2 &\leq \frac{2C}{n} \frac{1}{\lambda_{k_n}} 8k_n^2 \left( \sum_{l'=1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \right)^2 \sum_{l=k_n+1}^{\infty} \lambda_l \\ &\leq \frac{C}{n} k_n^2 \left( \sum_{l'=1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \right)^2 (k_n + 2) \frac{\lambda_{k_n+1}}{\lambda_{k_n}} \\ &\leq \frac{C}{n} k_n^3 \left( \sum_{l'=1}^{\infty} |\langle \boldsymbol{\rho}, \mathbf{e}_{l'} \rangle| \right)^2, \end{aligned}$$

where we have applied Lemma 1 in CMS. Obviously this quantity converges to zero because of **A4** and **A9**. This proves that  $\sqrt{n} \mathbb{E} [|\langle \mathcal{S}_n \boldsymbol{\rho}, \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}} \rangle|] \rightarrow 0$ , hence  $\sqrt{n} \langle \mathcal{S}_n \boldsymbol{\rho}, \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}} \rangle \xrightarrow{P} 0$ .

Therefore, we only need to show that  $\sqrt{n}\langle \mathcal{R}_n \boldsymbol{\rho}, \bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}} \rangle \xrightarrow{p} 0$ .

To deal with this term, we follow the proof in pages 350 and 351 in CMS. According to the arguments in those pages, it happens that we only need to show that

$$\sqrt{n} \sum_{j=1}^{k_n} \int_{\mathcal{B}_j} \mathbb{E} \left[ \|G_n(z)\|_\infty^2 \|(zI - \Gamma)^{-1/2}(\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}})\| \right] \|\Gamma_z^{-1/2} \boldsymbol{\rho}\| dz \rightarrow 0. \quad (24)$$

According to Lemma 3 in CMS, if  $z \in \mathcal{B}_j$ , then

$$\|G_n(z)\|_\infty^2 \leq 4 \sum_{l=1}^{\infty} \sum_{\substack{k=1 \\ k \neq j}}^{\infty} \frac{\langle (\Gamma_n - \Gamma) \mathbf{e}_l, \mathbf{e}_k \rangle^2}{|\lambda_j - \lambda_l| |\lambda_j - \lambda_k|} + 2 \sum_{\substack{k=1 \\ k \neq j}}^{\infty} \frac{\langle (\Gamma_n - \Gamma) \mathbf{e}_j, \mathbf{e}_k \rangle^2}{\delta_j |\lambda_j - \lambda_k|} + \frac{\langle (\Gamma_n - \Gamma) \mathbf{e}_j, \mathbf{e}_j \rangle^2}{\delta_j^2}.$$

From (16) we have that, if we take the positive value of the square root, then

$$\|\Gamma_z^{-1/2}(\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}})\| \leq \frac{1}{n} \sum_{i=1}^{\infty} \sqrt{\frac{\lambda_i}{|z - \lambda_i|}} \sum_{r=1}^n D_{x,\mathbf{h}}^{r,i}.$$

Thus, if  $z \in \mathcal{B}_j$ , then

$$\begin{aligned} \mathbb{E} \left[ \|G_n(z)\|_\infty^2 \|\Gamma_z^{-1/2}(\bar{\mathbf{X}}_{x,\mathbf{h}} - \mathbf{E}_{x,\mathbf{h}})\| \right] &\leq 4 \sum_{i=1}^{\infty} \sum_{l=1}^{\infty} \sum_{\substack{k=1 \\ k \neq j}}^{\infty} \sqrt{\frac{\lambda_i}{|z - \lambda_i|}} \frac{\mathbb{E} \left[ D_{x,\mathbf{h}}^{1,i} \langle (\Gamma_n - \Gamma) \mathbf{e}_l, \mathbf{e}_k \rangle^2 \right]}{|\lambda_j - \lambda_l| |\lambda_j - \lambda_k|} \\ &\quad + 2 \sum_{i=1}^{\infty} \sum_{\substack{k=1 \\ k \neq j}}^{\infty} \sqrt{\frac{\lambda_i}{|z - \lambda_i|}} \frac{\mathbb{E} \left[ D_{x,\mathbf{h}}^{1,i} \langle (\Gamma_n - \Gamma) \mathbf{e}_j, \mathbf{e}_k \rangle^2 \right]}{\delta_j |\lambda_j - \lambda_k|} \\ &\quad + \sum_{i=1}^{\infty} \sqrt{\frac{\lambda_i}{|z - \lambda_i|}} \frac{\mathbb{E} \left[ D_{x,\mathbf{h}}^{1,i} \langle (\Gamma_n - \Gamma) \mathbf{e}_j, \mathbf{e}_j \rangle^2 \right]}{\delta_j^2}. \end{aligned} \quad (25)$$

It is not difficult to check that  $\langle (\Gamma_n - \Gamma) \mathbf{e}_l, \mathbf{e}_k \rangle = \frac{\sqrt{\lambda_l \lambda_k}}{n} \sum_{s=1}^n (\xi_l^s \xi_k^s - \delta_{l,k})$ , so

$$D_{x,\mathbf{h}}^{1,i} \langle (\Gamma_n - \Gamma) \mathbf{e}_l, \mathbf{e}_k \rangle^2 = \frac{\lambda_l \lambda_k}{n^2} \sum_{s,s'=1}^n D_{x,\mathbf{h}}^{1,i} (\xi_l^s \xi_k^s - \delta_{l,k}) (\xi_l^{s'} \xi_k^{s'} - \delta_{l,k}).$$

Let  $s, s' \in \{1, \dots, n\}$ . If  $s \neq s'$ , at least one of them is different from 1. Assume that it happens that  $s \neq 1$ . By the independence of the sample, we have that

$$\mathbb{E} \left[ D_{x,\mathbf{h}}^{1,i} (\xi_l^s \xi_k^s - \delta_{l,k}) (\xi_l^{s'} \xi_k^{s'} - \delta_{l,k}) \right] = \mathbb{E} [(\xi_l^s \xi_k^s - \delta_{l,k})] \mathbb{E} [D_{x,\mathbf{h}}^{1,i} (\xi_l^{s'} \xi_k^{s'} - \delta_{l,k})] = 0.$$

Similarly, if  $s = s'$  and  $s \neq 1$ , we have  $\mathbb{E} [D_{x,\mathbf{h}}^{1,i} (\xi_l^s \xi_k^s - \delta_{l,k}) (\xi_l^{s'} \xi_k^{s'} - \delta_{l,k})] = 0$ . Thus,

$$\mathbb{E} \left[ D_{x,\mathbf{h}}^{1,i} \langle (\Gamma_n - \Gamma) \mathbf{e}_l, \mathbf{e}_k \rangle^2 \right] = \frac{\lambda_l \lambda_k}{n^2} \mathbb{E} \left[ D_{x,\mathbf{h}}^{1,i} (\xi_l^1 \xi_k^1 - \delta_{l,k})^2 \right].$$

On the other hand, it can be checked that, for every  $i, l, k$ , it happens that  $|\mathbb{E} [D_{x,\mathbf{h}}^{1,i} (\xi_l^1 \xi_k^1 - \delta_{l,k})^2]| \leq CM$ , where  $M$  is given in **A8**. For instance, let us assume that  $l = k$  with  $k \neq i$ . We have that

$$\left| \mathbb{E} \left[ D_{x,\mathbf{h}}^{1,i} (\xi_l^1 \xi_k^1 - \delta_{l,k})^2 \right] \right| \leq \mathbb{E} [|\xi_i^1| (|\xi_k^1|^2 + 1)^2] + \mathbb{E} [\mathbb{E} [|\xi_i^1|] (|\xi_k^1|^2 + 1)^2].$$

The term with higher order expectations is  $\mathbb{E} [|\xi_i^1| |\xi_k^1|^4]$  and it can be bounded:

$$\mathbb{E} [|\xi_i^1| |\xi_k^1|^4] \leq \mathbb{E} [\max(|\xi_i^1|, |\xi_k^1|)^5] \leq \mathbb{E} [|\xi_i^1|^5 + |\xi_k^1|^5] \leq 2M.$$

In summary, we have that

$$\left| \mathbb{E} \left[ D_{x, \mathbf{h}}^{1,i} \langle (\Gamma_n - \Gamma) \mathbf{e}_l, \mathbf{e}_k \rangle^2 \right] \right| \leq \frac{C \lambda_l \lambda_k}{n^2}$$

From here and (25) we obtain that, if  $z \in \mathcal{B}_j$ , then

$$\begin{aligned} & \mathbb{E} \left[ \|G_n(z)\|_\infty^2 \|\Gamma_z^{-1/2}(\bar{\mathbf{X}}_{x, \mathbf{h}} - \mathbf{E}_{x, \mathbf{h}})\| \right] \\ & \leq \frac{C}{n^2} \left( \sum_{l=1}^{\infty} \sum_{\substack{k=1 \\ k \neq j}}^{\infty} \frac{\lambda_l \lambda_k}{|\lambda_j - \lambda_l| |\lambda_j - \lambda_k|} + \sum_{\substack{k=1 \\ k \neq j}}^{\infty} \frac{\lambda_j \lambda_k}{\delta_j |\lambda_j - \lambda_k|} + \frac{\lambda_j^2}{\delta_j^2} \right) \sum_{i=1}^{\infty} \sqrt{\frac{\lambda_i}{|z - \lambda_i|}} \\ & \leq \frac{C}{n^2} (j \log j)^2 \left( \sum_{\substack{i=1 \\ i \neq j}}^{\infty} \sqrt{\frac{\lambda_i}{|\lambda_j - \lambda_i|}} + \sqrt{\frac{\lambda_j}{\delta_j}} \right). \end{aligned} \quad (26)$$

where we have applied the same argument as in the final part of the proof of Lemma 3 in CMS and Lemma A.3 in our paper.

Since  $\delta_j \geq \lambda_j - \lambda_{j+1}$ , first part in Lemma 1 in CMS gives that

$$\sqrt{\frac{\lambda_j}{\delta_j}} \leq \sqrt{\frac{\lambda_j}{\lambda_j - \lambda_{j+1}}} \leq \sqrt{j+1} \leq 2\sqrt{k_n}. \quad (27)$$

On the other hand, we have that

$$\sum_{\substack{i=1 \\ i \neq j}}^{\infty} \sqrt{\frac{\lambda_i}{|\lambda_j - \lambda_i|}} \leq \sqrt{\frac{\lambda_j}{\delta_j}} \frac{1}{\sqrt{\lambda_j}} \sum_{\substack{i=1 \\ i \neq j}}^{\infty} \sqrt{\lambda_i} \leq C \sqrt{k_n} \frac{1}{\sqrt{c_n}}, \quad (28)$$

where we have employed that  $\delta_j \leq |\lambda_j - \lambda_i|$  if  $i \neq j$ , (27), that **A6** implies  $\sum_{\substack{i=1 \\ i \neq j}}^{\infty} \sqrt{\lambda_i} < \infty$  and that, by the definitions of  $\delta_j$  and  $k_n$ , we have that for every  $j \leq k_n$ ,  $c_n \leq \lambda_j + \frac{\delta_j}{2} \leq 3\frac{\lambda_j}{2}$ . Inequalities (26), (27) and (28) give that

$$\begin{aligned} & \int_{\mathcal{B}_j} \mathbb{E} \left[ \|G_n(z)\|_\infty^2 \|(zI - \Gamma)^{-1/2}(\bar{\mathbf{X}}_{x, \mathbf{h}} - \mathbf{E}_{x, \mathbf{h}})\| \right] \|\Gamma_z^{-1/2} \boldsymbol{\rho}\| dz \\ & \leq C \text{diam}(\mathcal{B}_j) \frac{(j \log j)^2 \sqrt{k_n}}{n^2 \sqrt{c_n}} \sup_{z \in \mathcal{B}_j} \|\Gamma_z^{-1/2} \boldsymbol{\rho}\| \\ & \leq C \delta_j \frac{(j \log j)^2 \sqrt{k_n}}{n^2 \sqrt{c_n}} \frac{1}{\sqrt{\delta_j}} \|\boldsymbol{\rho}\| \leq C \frac{(j \log j)^{3/2} \sqrt{k_n}}{n^{7/4}}, \end{aligned}$$

where we have applied the facts that if  $z \in \mathcal{B}_j$  then  $|z - \lambda_i| \geq \delta_j/2$ , that if  $j$  is large enough then  $\delta_j < C(j \log j)^{-1}$  and **A9**. From here, we have that

$$\sqrt{n} \sum_{j=1}^{k_n} \int_{\mathcal{B}_j} \mathbb{E} \left[ \|G_n(z)\|_\infty^2 \|(zI - \Gamma)^{-1/2}(\bar{\mathbf{X}}_{x, \mathbf{h}} - \mathbf{E}_{x, \mathbf{h}})\| \right] \|\Gamma_z^{-1/2} \boldsymbol{\rho}\| dz \leq C \frac{k_n^3 (\log k_n)^{3/2}}{n^{5/4}} \rightarrow 0,$$

by Lemma A.1. This proves (24) and, consequently, that  $n^{1/2} \langle \bar{\mathbf{X}}_{x, \mathbf{h}} - \mathbf{E}_{x, \mathbf{h}}, \mathbf{Y}_n \rangle = o_{\mathbb{P}}(1)$ .  $\square$

*Proof of Lemma A.7.* According to Lemma 8, page 355, in CMS, we have that

$$\frac{\sqrt{n}}{t_{n,\mathbf{E}_{x,\mathbf{h}}}} \langle \mathbf{E}_{x,\mathbf{h}}, \mathbf{R}_n \rangle \overset{\mathcal{L}}{\rightsquigarrow} \mathcal{N}(0, \sigma_\varepsilon^2).$$

From the proof of Proposition 3.2 we have that  $\sup_j \frac{\langle \mathbf{E}_{x,\mathbf{h}}, \mathbf{e}_j \rangle^2}{\lambda_j} \leq 1$ . Since the sequence  $\{t_{n,\mathbf{E}_{x,\mathbf{h}}}\}_n$  is strictly increasing, with its terms strictly positive, Lemma A.1 implies that  $\frac{k_n^3 (\log k_n)^2}{\sqrt{n} t_{n,\mathbf{E}_{x,\mathbf{h}}}} \rightarrow 0$ . Therefore, the second part of Proposition 3 in CMS (page 352) gives that  $\langle \mathbf{E}_{x,\mathbf{h}}, \mathbf{Y}_n \rangle = o_{\mathbb{P}}(n^{-1/2})$ .

Then, the result will be proved if we show that

$$n^{1/2} (\langle \mathbf{E}_{x,\mathbf{h}}, \mathbf{L}_n \rangle + \langle \mathbf{E}_{x,\mathbf{h}}, \mathbf{S}_n \rangle) = o_{\mathbb{P}}(1). \quad (29)$$

To prove (29), we analyse separately both two terms in the left hand side of this expression. To this, we follow the steps in some proofs in CMS. Before, notice that in the computation of  $\mathbf{E}_{x,\mathbf{h}}$  we can take  $\mathbf{X}$  independent from all the remaining variables in the problem. In particular, for every  $n \in \mathbb{N}$ , we can consider that  $\mathbf{X}$  is independent of  $\mathbf{W}_n$  where  $\mathbf{W}_n$  is either  $\mathbf{L}_n$  or  $\mathbf{S}_n$ . Thus, we have that,

$$\begin{aligned} \mathbb{E} \left[ \left| \left\langle \mathbb{E} \left[ \mathbb{1}_{\{\mathbf{X}^h \leq x\}} \mathbf{X} \right], \mathbf{W}_n \right\rangle \right| \right] &= \mathbb{E} \left[ \left| \mathbb{E} \left[ \left\langle \mathbb{1}_{\{\mathbf{X}^h \leq x\}} \mathbf{X}, \mathbf{W}_n \right\rangle \middle| \mathbf{W}_n \right] \right| \right] \\ &\leq \mathbb{E} \left[ \mathbb{1}_{\{\mathbf{X}^h \leq x\}} |\langle \mathbf{X}, \mathbf{W}_n \rangle| \right] \leq \mathbb{E} [|\langle \mathbf{X}, \mathbf{W}_n \rangle|]. \end{aligned}$$

Hence, if we show  $\mathbb{E} [|\langle \mathbf{X}, \mathbf{W}_n \rangle|] = o_{\mathbb{P}}(n^{-1/2})$ , then Markov's inequality gives (29).

Lemma 7, page 345, in CMS gives the following two bounds:

$$\mathbb{E} [|\langle \mathbf{X}, \mathbf{L}_n \rangle|] \leq \begin{cases} |\langle \boldsymbol{\rho}, \mathbf{e}_{k_n} \rangle| \sqrt{\sum_{j=k_n+1}^{\infty} \lambda_j}, \\ \frac{\lambda_{k_n}}{\sqrt{k_n \log k_n}} \sqrt{\sum_{j=k_n+1}^{\infty} \langle \boldsymbol{\rho}, \mathbf{e}_j \rangle}. \end{cases}$$

This inequality plus either **A3** or **A4** and **A7** yields  $\mathbb{E} [|\langle \mathbf{X}, \mathbf{L}_n \rangle|] = o_{\mathbb{P}}(n^{-1/2})$ .

To handle the term  $n^{1/2} \mathbb{E} [|\langle \mathbf{X}, \mathbf{S}_n \rangle|]$  we follow the argument in Proposition 2, page 346, in CMS. We have that

$$\langle \mathbf{X}, \mathbf{S}_n \rangle = \langle \mathbf{X}, \mathcal{R}_n \boldsymbol{\rho} \rangle + \langle \mathbf{X}, \mathcal{S}_n \boldsymbol{\rho} \rangle.$$

Lemma A.1 implies that  $k_n^2 \log k_n = o(n^{1/2})$  and, then, the reasoning in the last part of the proof of Proposition 2, page 352, in CMS leads to

$$\mathbb{E} [|\langle \mathcal{R}_n \boldsymbol{\rho}, \mathbf{X} \rangle|] \leq C \frac{k_n^3 (\log k_n)^2}{n} = o(n^{-1/2}).$$

Concerning the term  $\langle \mathbf{X}, \mathcal{S}_n \boldsymbol{\rho} \rangle$ , let us consider the decomposition of  $\mathbb{E} [\langle \mathbf{X}, \mathcal{S}_n \boldsymbol{\rho} \rangle^2]$ , which appears in (24) and (25) in CMS. The term in (24) is bounded by the expression in (26) and then the authors of CMS find a bound for each term in (26). The difference with our case is that our term  $\mathbb{E} [\langle \mathbf{X}, \mathcal{S}_n \boldsymbol{\rho} \rangle]^2$  is not divided over  $k_n$ . Therefore, in order to be able to apply the inequality in the second display in page 350 in CMS (which allow them to bound the second term in (26) of their paper) we need to reinforce the assumption  $\lambda_{k_n} k_n \log k_n \rightarrow 0$ , used in CMS, to  $\lambda_{k_n} k_n^3 \log k_n \rightarrow 0$ , but this follows from **A6**.

The bound for the first term in (26) of CMS in our case would be

$$k_n M \frac{k_n}{\log k_n} \lambda_{k_n} k_n^2 \max_{k_n < j \leq k_n + h_n} \{ \langle \boldsymbol{\rho}, \mathbf{e}_j \rangle^2 \}$$

which converges to zero by **A4** and **A6**. The convergence to zero of the term in (25) in CMS is proved similarly.  $\square$

## C Supplement to the simulation study

### C.1 Detailed description of simulation scenarios

The description of the simulation scenarios is given in Table 2. The functional processes, all of them indexed in  $[0, 1]$  and discretized in 201 equidistant points, are the following:

**BM.** Brownian motion, denoted by  $\mathbf{B}$ , whose eigenfunctions are  $\psi_j(t) = \sqrt{2} \sin((j - \frac{1}{2})\pi t)$ ,  $j \geq 1$ .

**HHN.** Functional process considered in Hall and Hosseini-Nasab (2006), given by  $\mathbf{X}(t) = \sum_{j=1}^{20} \xi_j \phi_j(t)$ , where  $\phi_j(t) = \sqrt{2} \cos(j\pi t)$  and  $\xi_j$  are independent r.v.'s distributed as  $\mathcal{N}(0, j^{-2l})$ , with  $l = 1, 2$ .

**BB.** Brownian bridge, defined as  $\mathbf{X}(t) = \mathbf{B}(t) - t\mathbf{B}(1)$ . Its eigenfunctions are  $\tilde{\psi}_j(t) = \psi_{j+\frac{1}{2}}(t)$ ,  $j \geq 1$ .

**OU.** Ornstein-Uhlenbeck process, defined as the zero-mean Gaussian process with covariance given by  $\text{Cov}[\mathbf{X}(s), \mathbf{X}(t)] = \frac{\sigma^2}{2\alpha} e^{-\alpha(s+t)} (e^{2\alpha \min(s,t)} - 1)$ . We consider  $\alpha = \frac{1}{3}$ ,  $\sigma = 1$  and  $\mathbf{X}(0) \sim \mathcal{N}(0, \frac{\sigma^2}{2\alpha})$ .

**GBM.** Geometric Brownian motion, defined as  $\mathbf{X}(t) = s_0 \exp\{(\mu - \frac{\sigma^2}{2})t + \sigma \mathbf{B}(t)\}$ . We consider  $\sigma = 1$ ,  $\mu = \frac{1}{2}$  and  $s_0 = 2$ .

Scenario	Coefficient $\boldsymbol{\rho}(t)$	Process $\mathbf{X}$	Deviation
S1	$(2\psi_1(t) + 4\psi_2(t) + 5\psi_3(t))/\sqrt{2}$	BM	$\delta_k \Delta_1$ , $\delta = (0, \frac{1}{4}, \frac{3}{4})$
S2	$(2\tilde{\psi}_1(t) + 4\tilde{\psi}_2(t) + 5\tilde{\psi}_3(t))/\sqrt{2}$	BB	$-\delta_k \Delta_2$ , $\delta = (0, 2, \frac{15}{2})$
S3	$(2\psi_2(t) + 4\psi_3(t) + 5\psi_7(t))/\sqrt{2}$	BM	$-\delta_k \Delta_1$ , $\delta = (0, \frac{1}{5}, \frac{1}{2})$
S4	$\sum_{j=1}^{20} 2^{3/2} (-1)^j j^{-2} \phi_j(t)$	HHN ( $l = 1$ )	$-\delta_k \Delta_2$ , $\delta = (0, 1, 3)$
S5	$\sum_{j=1}^{20} 2^{3/2} (-1)^j j^{-2} \phi_j(t)$	HHN ( $l = 2$ )	$-\delta_k \Delta_2$ , $\delta = (0, 1, 3)$
S6	$\log(15t^2 + 10) + \cos(4\pi t)$	BM	$\delta_k \Delta_1$ , $\delta = (0, \frac{1}{5}, 1)$
S7	$\sin(2\pi t) - \cos(2\pi t)$	OU	$-\delta_k \Delta_2$ , $\delta = (0, \frac{1}{4}, 1)$
S8	$t - (t - \frac{3}{4})^2$	OU	$-\delta_k \Delta_3$ , $\delta = (0, \frac{1}{100}, \frac{1}{10})$
S9	$\pi^2 (t^2 - \frac{1}{3})$	GBM	$\delta_k \Delta_3$ , $\delta = (0, \frac{1}{2}, \frac{5}{2})$

Table 2: Simulation scenarios and deviations from the null hypothesis.

The first scenario, S1, contains a  $\boldsymbol{\rho}$  based on example (a) in Section 5 of Cardot et al. (2003), which is a linear combination of the three first eigenfunctions of the Brownian motion. Variations of the same idea – a  $\boldsymbol{\rho}$  that is a finite linear combination of the eigenfunctions of the functional process – is employed in the next four scenarios. S2 considers a Brownian bridge as the functional process. S3 takes linear combinations of eigenfunctions associated to smaller eigenvalues to construct  $\boldsymbol{\rho}$ . S4 and S5 collect the process and coefficients used in Section 5 of Hall and Hosseini-Nasab (2006) for  $l = 1$  and  $l = 2$ , respectively, which is a finite-dimensional smooth process. S7 is example (b) in Cardot et al. (2003), which is not expressible as a finite combination of eigenfunctions. The remaining scenarios follow this idea: S7 and S8 with Ornstein-Uhlenbeck processes (as in Section 4.2 of García-Portugués et al. (2014)) and S9 with a geometric Brownian motion.



The deviation coefficients  $\delta_d$ ,  $d = 1, 2$ , for  $k = 1, \dots, 9$  were chosen by comparing the densities of the response  $Y$  under the null and the alternative. This comparison provides a graphical visualization of the hardness of distinguishing between both hypotheses (Figure 5). The actual choice of the coefficients aims to make this distinction a challenging task.

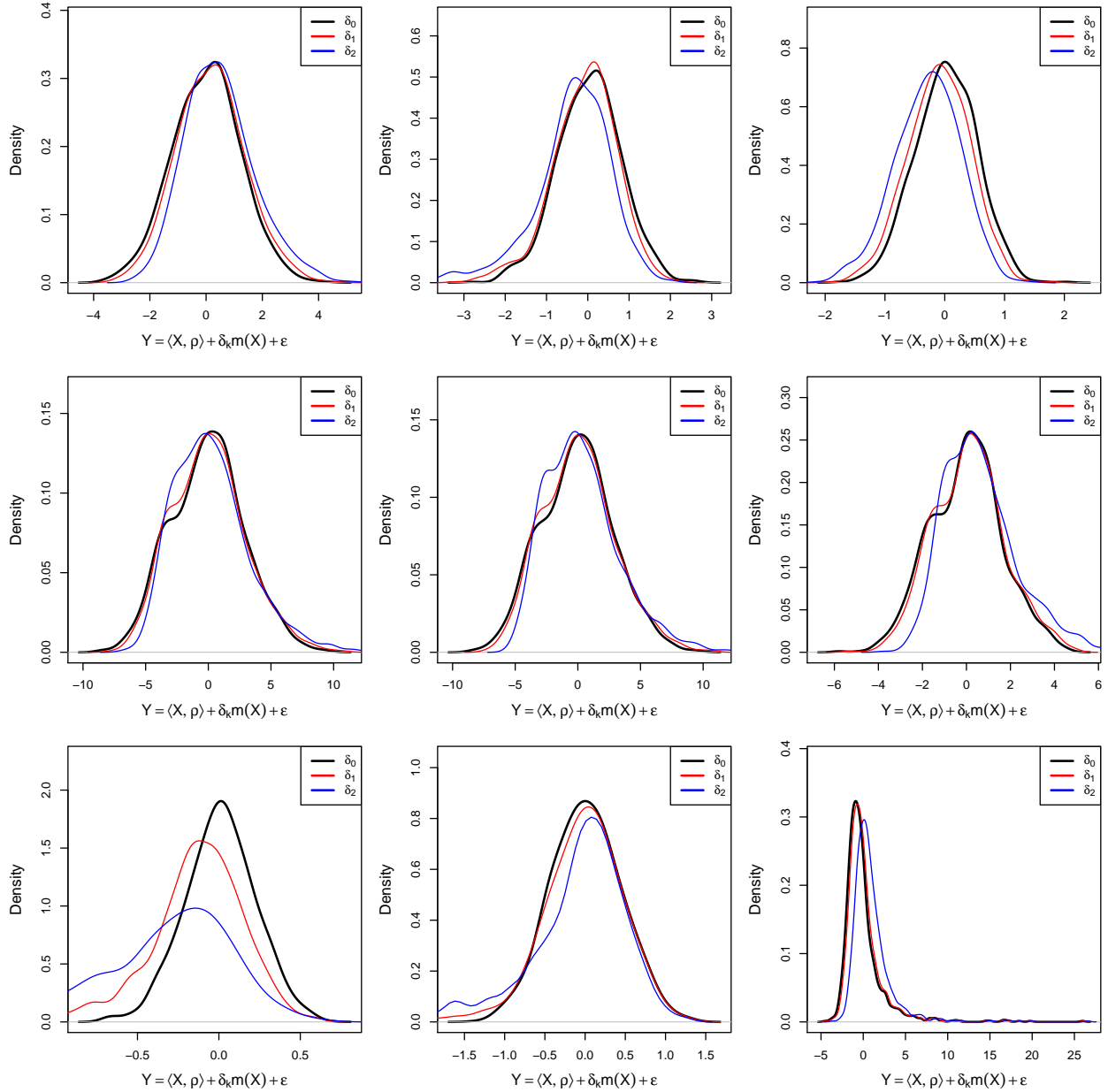


Figure 5: Densities of the responses  $Y$  under  $H_{k,d}$ , with  $k = 1, \dots, 9$  from left to right and up to down and  $d = 0, 1, 2$ .

## C.2 Supplementary results for the composite hypothesis

Table 3 shows the empirical sizes and powers for  $n = 50$ . The numerical summaries for the data-driven  $d_n$  are given in Table 4.

$H_{k,\delta}$	$n = 50$						
	CvM <sub>1</sub>	CvM <sub>5</sub>	CvM <sub>10</sub>	KS <sub>1</sub>	KS <sub>5</sub>	KS <sub>10</sub>	PCvM
$H_{1,0}$	0.060	0.046	0.039	0.056	0.042	0.041	0.046
$H_{2,0}$	0.082	0.071	0.061	0.088	0.077	0.067	0.078
$H_{3,0}$	0.061	0.035	0.030	0.069	0.041	0.033	0.056
$H_{4,0}$	0.067	0.047	0.036	0.069	0.054	0.034	0.041
$H_{5,0}$	0.055	0.035	0.033	0.056	0.041	0.031	0.051
$H_{6,0}$	0.056	0.041	0.038	0.059	0.045	0.045	0.031
$H_{7,0}$	0.074	0.048	0.038	0.075	0.060	0.053	0.053
$H_{8,0}$	0.046	0.033	0.032	0.045	0.037	0.032	0.042
$H_{9,0}$	0.077	0.058	0.056	0.071	0.063	0.048	0.072
$H_{1,1}$	0.309	0.270	0.261	0.228	0.220	0.210	0.359
$H_{2,1}$	0.671	0.696	0.681	0.541	0.573	0.552	0.788
$H_{3,1}$	0.150	0.113	0.104	0.130	0.107	0.089	0.155
$H_{4,1}$	0.188	0.184	0.167	0.146	0.139	0.134	0.216
$H_{5,1}$	0.255	0.202	0.202	0.179	0.161	0.149	0.251
$H_{6,1}$	0.652	0.814	0.801	0.552	0.696	0.674	0.857
$H_{7,1}$	0.940	0.988	0.988	0.913	0.964	0.959	0.991
$H_{8,1}$	0.557	0.550	0.544	0.347	0.349	0.342	0.582
$H_{9,1}$	0.102	0.071	0.067	0.100	0.073	0.070	0.081
$H_{1,2}$	0.883	0.963	0.959	0.804	0.889	0.887	0.985
$H_{2,2}$	0.918	0.988	0.985	0.877	0.945	0.944	0.996
$H_{3,2}$	0.852	0.950	0.949	0.776	0.867	0.869	0.976
$H_{4,2}$	0.545	0.695	0.690	0.458	0.540	0.546	0.778
$H_{5,2}$	0.827	0.853	0.837	0.701	0.727	0.713	0.889
$H_{6,2}$	0.779	0.967	0.970	0.741	0.899	0.912	0.976
$H_{7,2}$	0.959	0.997	0.996	0.936	0.971	0.970	0.996
$H_{8,2}$	0.666	0.676	0.674	0.343	0.344	0.349	0.712
$H_{9,2}$	0.514	0.546	0.538	0.454	0.490	0.478	0.640

Table 3: Empirical sizes and powers of the CvM, KS and PCvM tests with  $\alpha = 0.05$ , sample size  $n = 50$  and estimation of  $\rho$  by data-driven FPC ( $d_n$  chosen by SICc). KS and CvM tests are shown with 1, 5 and 10 projections.

$H_{k,\delta}$	$n = 50$			$n = 100$			$n = 250$		
	$\delta = 0$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 1$	$\delta = 2$
$H_{1,\delta}$	3.40 (0.64)	3.39 (0.65)	3.28 (0.57)	3.35 (0.61)	3.35 (0.58)	3.23 (0.54)	3.30 (0.52)	3.29 (0.56)	3.18 (0.46)
$H_{2,\delta}$	3.69 (0.83)	3.50 (0.78)	3.02 (0.61)	3.75 (0.84)	3.46 (0.68)	3.10 (0.38)	3.61 (0.73)	3.39 (0.64)	3.06 (0.25)
$H_{3,\delta}$	7.90 (1.15)	7.60 (1.26)	6.30 (1.92)	8.16 (1.05)	7.99 (1.04)	7.52 (0.82)	8.12 (1.04)	7.92 (0.93)	7.56 (0.73)
$H_{4,\delta}$	2.05 (0.69)	2.00 (0.68)	1.79 (0.72)	2.29 (0.59)	2.29 (0.65)	2.03 (0.61)	2.62 (0.68)	2.56 (0.63)	2.33 (0.57)
$H_{5,\delta}$	1.45 (0.58)	1.42 (0.59)	1.34 (0.57)	1.75 (0.58)	1.68 (0.58)	1.45 (0.56)	2.04 (0.35)	2.02 (0.37)	1.82 (0.49)
$H_{6,\delta}$	1.56 (0.69)	1.57 (0.74)	1.39 (0.65)	1.94 (0.80)	1.94 (0.77)	1.59 (0.70)	2.53 (0.99)	2.49 (0.93)	2.05 (0.76)
$H_{7,\delta}$	4.14 (0.70)	3.15 (0.64)	1.55 (0.76)	4.28 (0.64)	3.34 (0.54)	1.82 (0.78)	4.37 (0.72)	3.65 (0.57)	2.36 (0.70)
$H_{8,\delta}$	2.13 (0.43)	1.90 (0.48)	1.19 (0.52)	2.18 (0.51)	1.97 (0.45)	1.12 (0.36)	2.29 (0.55)	1.98 (0.34)	1.11 (0.33)
$H_{9,\delta}$	3.02 (0.95)	2.99 (0.96)	2.76 (0.89)	3.49 (1.09)	3.44 (1.07)	3.22 (1.05)	4.25 (1.02)	4.29 (1.02)	3.93 (0.98)

Table 4: Averages of the SICc-driven  $d_n$  for the different models, sample sizes and deviations from the null. Standard deviations are given between parentheses.

### C.3 Dependence on the projection process

The goodness-of-fit tests depend clearly on the way random directions  $\mathbf{h}$  are chosen. In order to explore its practical influence, we replicated the results in Section 5 for two different processes: the first, denoted  $(ii)$ , is based on the data-driven process described in Section 4 – which is referred as  $(i)$  – but with constant coefficients  $\eta_j \sim \mathcal{N}(0, 1)$ ; the second,  $(iii)$ , is an Ornstein-Uhlenbeck process with  $\alpha = \frac{1}{2}$  and  $\sigma = 1$  that is completely independent from the sample. Process  $(ii)$  generates more noisy random directions, since all the FPCs of the sample are equally weighted.

Figures 6 and 7 show the empirical levels and powers of the tests based on processes  $(i)$ ,  $(ii)$  and  $(iii)$ , for  $n = 100$ . Relatively minor changes can be observed between  $(i)$  and  $(iii)$ , with the main features described in Section 5 being consistent: L-shaped patterns in the size curves, mild decrements for the power curves, occasional bumps yielding power gains, and domination of CvM over KS. The results for both processes show no main changes, both indicating that  $K \in [1, 10]$  is a reasonable choice with respect to size and power. The big picture for  $(ii)$  is similar, albeit with more spread and variable level curves, and power curves dominated by the ones of  $(i)$  and  $(iii)$ .

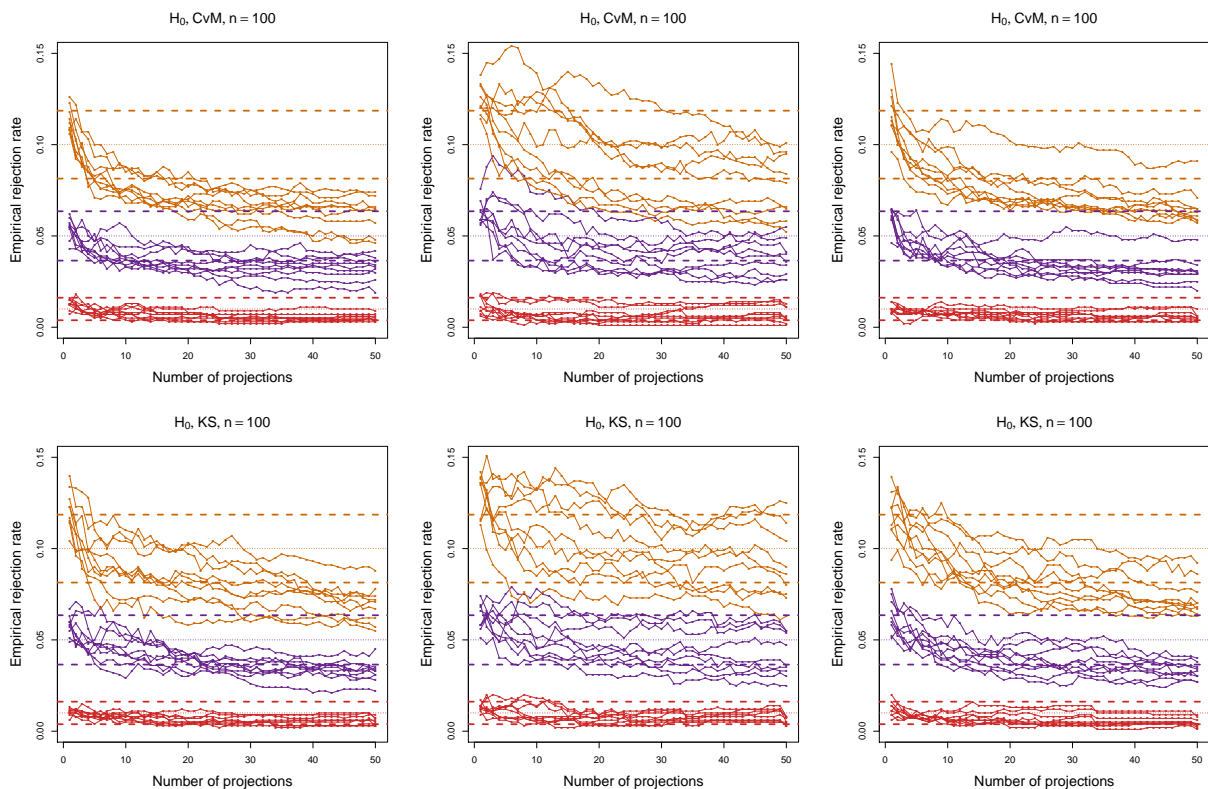


Figure 6: Empirical sizes of the CvM (upper row) and KS (lower row) tests for  $Sk$ ,  $k = 1, \dots, 9$ , depending on the number of projections  $K = 1, \dots, 50$ . From left to right, columns represent the data generating processes  $(i)$ ,  $(ii)$  and  $(iii)$ . The sample size is  $n = 100$ . The empirical sizes associated to the significance levels  $\alpha = 0.01, 0.05, 0.10$  are coded in red, purple and orange, respectively. Dashed thick lines represent the asymptotic 95% confidence interval for the proportion  $\alpha$  obtained from  $M$  replicates.

The presented empirical results indicate that less variable random directions seem to yield a better behaviour for the tests and that the data-driving process given in Section 4 is a sensible alternative. However, a more thorough research into the selection of the projecting process – out of the scope of the paper – is required.

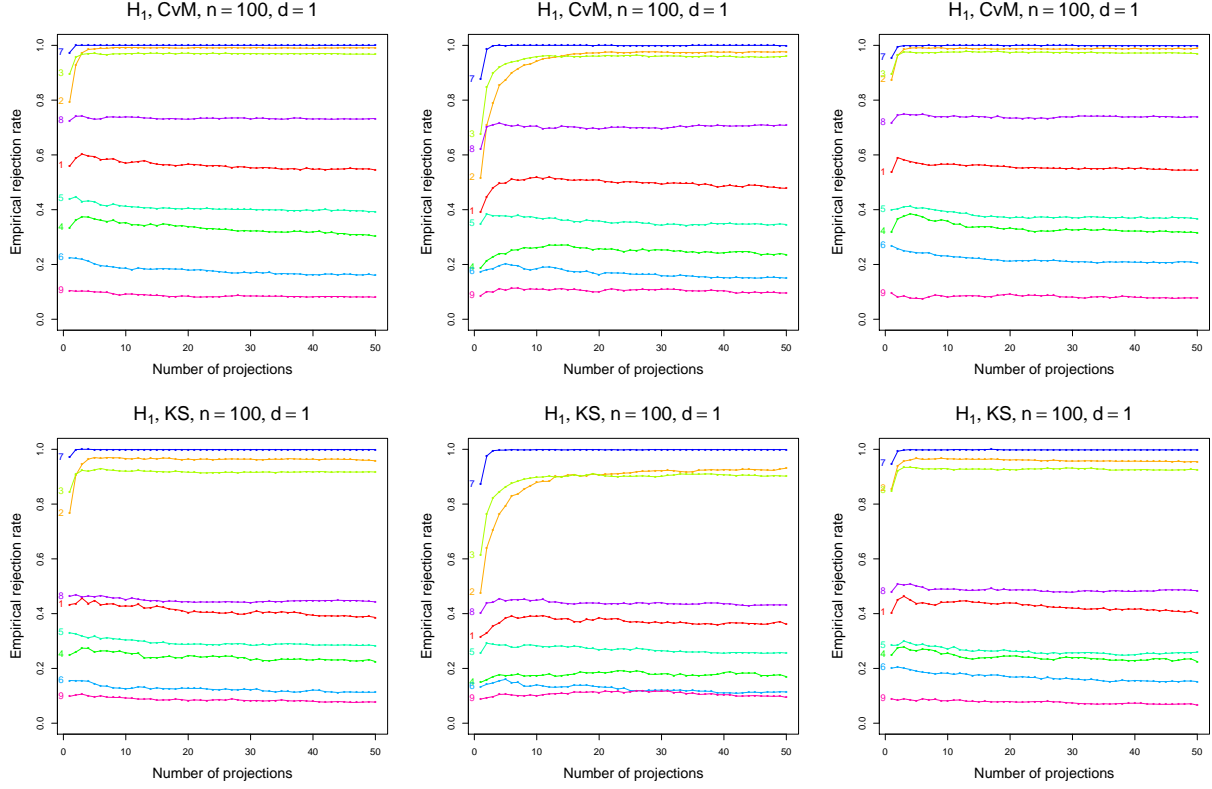


Figure 7: Empirical powers of the CvM (first row) and KS (second two) tests for  $Sk$ ,  $k = 1, \dots, 9$ , depending on the number of projections  $K = 1, \dots, 50$ . From left to right, columns represent the data generating processes (i), (ii) and (iii). The significance level is  $\alpha = 0.05$ , the sample size is  $n = 100$  and the deviation index is  $d = 1$ .

#### C.4 Remarks on discrete $p$ -values and FDR correction

The FDR correction of a set of  $K$  continuous  $p$ -values  $p_1, \dots, p_K$ , arising from  $K$  hypothesis tests, results in the FDR  $p$ -value  $p_{\text{FDR},K} := \min_{i=1,\dots,K} \frac{K}{i} p_{(i)}$ . Under the null hypotheses for all the tests, the level of the test that rejects if  $p_{\text{FDR},K} < \alpha$  is  $\alpha$  at most (Benjamini and Yekutieli, 2001). When using a resampling strategy for approximating the  $p$ -values  $p_1, \dots, p_K$ , for example by considering  $B$  bootstrap replicates, we end up with a collection of discrete  $p$ -values  $\hat{p}_1, \dots, \hat{p}_K$ . This has a notable influence in  $p_{\text{FDR},K}$ , resulting in an increment of the type I error that is magnified for moderate and large  $K$ .

Under the null,  $\hat{p}_i$  is approximately distributed as a  $\mathcal{U}(\{0, \frac{1}{B}, \dots, \frac{B}{B}\})$ ,  $i = 1, \dots, K$ . If independence between  $\hat{p}_1, \dots, \hat{p}_K$  is assumed, then the rejection rate of  $p_{\text{FDR},K} < \alpha$  is at least  $\mathbb{P}[\hat{p}_{(1)} = 0] \approx 1 - \left(\frac{B}{B+1}\right)^K$  no matter what significance level  $\alpha$  is chosen. For example, if  $K = 25$  and  $B = 500$ , under the null  $p_{\text{FDR},K} < \alpha$  will reject at least 4.87% of the times. For  $K = 5$  and  $B = 1000$ , the lower bound for the rejection percentage drops to 0.499%. This simple argument illustrates the more demanding precision (larger  $B$ 's) required in the approximated  $p$ -values when  $K$  grows.

In order to gain more insights about the problematic dependence of  $K$  and  $B$ , we have conducted the following experiment, aimed to reproduce a comparable situation to our testing in practise.

1. Simulate  $K$  discrete  $p$ -values independently:  $\hat{p}_i \sim \mathcal{U}(\{0, \frac{1}{B}, \dots, \frac{B}{B}\})$  and compute  $p_{\text{FDR},K}$ .
2. Repeat the above steps  $M = 1000$  times and obtain the empirical rejection rates of  $p_{\text{FDR},K} < \alpha$  for  $\alpha = 0.10, 0.05, 0.01$ . Plot the rejection curves as a function of  $K = 1, \dots, 50$ . Repeat this

five times to account for variability.

3. Repeat the above steps for different  $B$ 's.

The results of the experiment, namely the empirical rejection curves as a function of  $K$  for different  $\alpha$ 's and  $B$ 's, are shown in the left panel of Figure 8. A sawtooth pattern of over-rejection appears for curves with  $B = 500, 1000$  (yellow and light green curves) and values of  $K$  larger than  $8 - 10$ , resulting in significant violations of the confidence intervals for the proportions  $\alpha$  for  $K$ 's in the range  $[10, 50]$ . When  $B$  is larger (dark green and blue curves), the rejection rates remain more stable and inside the confidence intervals for  $K$  up to 50. This highlights that, given the effect that both  $K$  and  $B$  have in the computation proficiency of the test, a reasonable compromise on the choice of  $K$  and  $B$  that respects the type I error is  $K \in [1, 10]$  and  $B = 1000$ . The right panel of Figure 8 shows the resulting levels if the positive correction  $\frac{\hat{p}B+1}{B+1}$  is applied for avoiding null  $p$ -values. The same conclusions can be extracted, the main change being under-rejections instead of over-rejections.

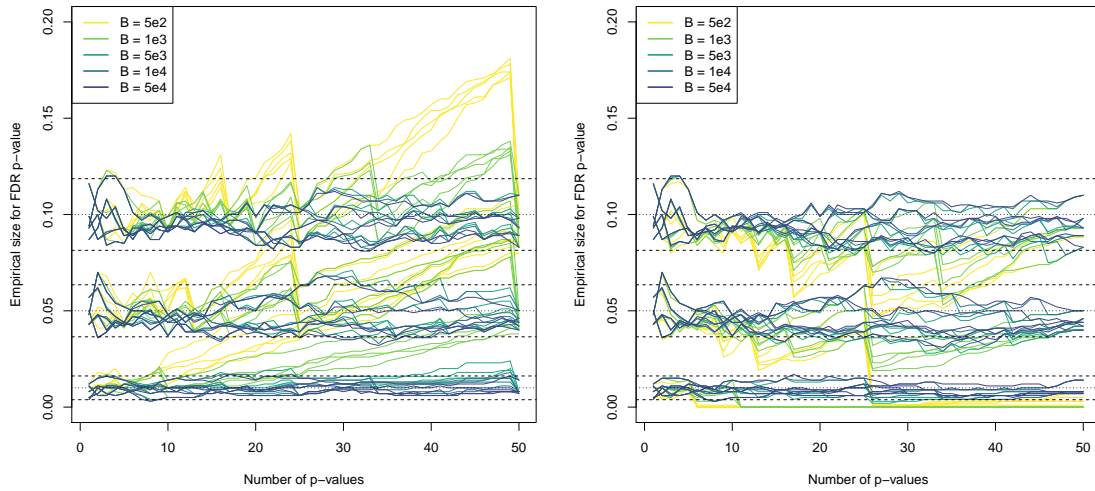


Figure 8: Left: empirical levels of the test with rejection rule  $p_{\text{FDR},K} < \alpha$ , as a function of  $K$ . Right: empirical levels employing a positive correction for the  $p$ -values.